Variable methylation of endogenous retroviruses: epigenetic inheritance, environmental modulation, and genetic modifiers

Tessa Mariah Bertozzi

Downing College

This dissertation is submitted for the degree of

Doctor of Philosophy

Department of Genetics

University of Cambridge

September 2019

To my parents.

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the text. The occasional use of 'we' is a reflection of my own work, unless specifically stated otherwise.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

It does not exceed the prescribed word limit for the Degree Committee of the School of the Biological Sciences.

Tessa Mariah Bertozzi

September 2019

Acknowledgements

I would like to thank my supervisor, Anne Ferguson-Smith, for all her support and guidance. Thank you for your unwavering enthusiasm for the project and for trusting so wholeheartedly in my ability to carry it out. You have nurtured and boosted my professional self-confidence more than anyone ever has.

I am incredibly grateful to all the current and former members of the AFS lab for their kindness and scientific input. It has been a privilege to work with such a wonderful group of scientists. Thank you to Anastasiya Kazachenka for so willingly bringing me on board to the metastable epiallele project; to Mitsu Ito, Nozomi Takahashi, and Geula Hanin for their patience and instruction on various lab techniques; to Joe Gardner, Célia Delahaye, and Rahia Mashoodh for making me feel welcome from day one; to Jenny Corish, Shrina Patel, and Fran Dearden for their help with mouse colony management; to Carol Edwards for reading my introduction; and to my students Jessica Elmer, Emily Simmons, Anna Olech, and Jesse Loilargosain for all their hard work. Special acknowledgements go to Jessica Elmer, Amir Hay, and Noah Kessler for their proofreading efforts, insightful discussions, and friendship; working and 'working' with you has been a blast.

Thank you to my adviser, Alfonso Martinez-Arias, for his useful feedback on my six week and first year reports. To Alwyn Scally for his joint supervision of Anna and for his advice on the modelling of stochastic methylation. To Felipe Karam Teixeira and Michael Imbeault for valuable scientific discussions and career mentorship. To Ali Robinson for taking such good care of my jumpy mice. To Matt Castle and Tony Fulford for guidance on statistical analyses. To Claire Wilson, Lottie Groocock, and Meg Staff for administrative support.

The scope of this project was significantly expanded by our many collaborators. I am grateful to Erica Watson, Gina Blake, Joan Barau, Déborah Bourch'is, Amita Bansal, Rebecca Simmons, Marisa Bartolomei, Duy Nguyen, Sue Ozanne, Denise Fernandez-Twinn, and Anne Corcoran for sharing samples from their respective mouse models. Thank you to Greg Cox, David Schroeder, and Jennifer Stauffer for their efforts on our A^{vy} collaboration. I am particularly appreciative of Erica Watson for her mentorship, availability, and attention to detail.

My PhD concluded with a thought-provoking conversation with my viva examiners, Carmen Sapienza and Paul Lehner. I thank them for their careful consideration of my work.

This research would not have been possible without generous financial support from the Cambridge Trust, Pomona College, Downing College, and the Cambridge Philosophical Society.

Thank you to Peggy Skinner, my high school biology teacher, for assigning that extra reading on *Agouti viable yellow* – who would have thought it would be the topic of my PhD so many years later.

I am grateful to all the new friends I have made in Cambridge and all the old ones who have kept in touch from afar. It would be difficult to list you all but know that your friendship has kept a smile on my face throughout.

I owe a great deal of gratitude to my family. Thank you to my loving parents for their endless encouragement and advice – I have you to thank for having come this far. Thank you to my siblings, Galen and Lena, for cheering me on the whole way. Thank you to my uncle and aunt, David and Jane, for being my home away from home and for letting me write a substantial portion of this dissertation in their house in Aldeburgh. Thank you to Nainie for showing such sincere interest in my work.

Finally, a heartfelt thank you to Tom for his unconditional love and support over the past three years. Now you know more about metastable epialleles than you ever needed (or wanted?) to know.

Summary

Half of the mammalian genome is made up of transposable elements (TEs), the vast majority of which are modified by DNA methylation and repressive histone marks. Intracisternal A-particles (IAPs) are evolutionarily young murine-specific endogenous retroviruses (ERVs). Some are capable of retrotransposition and many harbour regulatory sequences with the potential to influence host gene expression. Previous work has shown that, at the *Agouti Viable Yellow* (A^{vy}) locus, an IAP insertion provides an alternative promoter for the *Agouti* coat colour gene. Variable methylation of this IAP is correlated with coat colour expressivity. In addition, the A^{vy} allele exhibits transgenerational epigenetic inheritance, susceptibility to environmental exposures, and strain-specific modulation.

A recent screen conducted in the C57BL/6J mouse strain identified a subset of IAPs that show variable methylation levels across genetically identical individuals. This thesis characterises this novel repertoire of variably methylated IAPs (VM-IAPs) and in doing so assesses the extent to which endogenous retroviruses in the mammalian genome display A^{vy}-like properties. Findings indicate that VM-IAPs become hypermethylated in the male germline by the DNA methyltransferase DNMT3C and are reconstructed as variable loci in the next generation. An exploration of inter-individual VM-IAP methylation levels in the C57BL/6J population demonstrates that their frequency distributions form skewed bell curves that are locus-specific. Only one out of six VM-IAPs analysed via linear mixed effects models shows memory of maternal methylation level,

and the effect size is small. This challenges the generalisability of epigenetic inheritance at these regions.

In studying environmental influences on VM-IAPs, results suggest that these epigenetically dynamic loci are selectively susceptible to altered environmental conditions: abnormal folate metabolism appears to shift VM-IAP methylation levels and influence VM-IAP-associated gene expression, while no significant effects are observed following exposure to the endocrine-disruptor Bisphenol-A (BPA) or to an obesogenic diet. A longitudinal ageing analysis indicates that VM-IAPs are stable across the murine lifespan, with only modest increases in methylation detected for a subset of loci. Reciprocal hybrid breeding experiments reveal that VM-IAP methylation levels are subject to maternal and genetic background effects that can be harnessed to map strain-specific VM-IAP modifiers. Finally, a naturally occurring genetically-conferred epiallele that influences neighbouring gene expression is identified and characterised.

This body of work highlights the value of using VM-IAPs as a model to study TE regulation and mammalian epigenetic stochasticity, and serves as a foundation to elucidate the consequences of partially methylated repeat elements on genome structure and function.

Contents

List of Figures

List of Tables

Abbreviations

5hmC..... 5-hydroxymethylcytosine
5mC..... 5-methylcytosine
A..... Adenine
B6..... C57BL/6J
BC..... C57BL/6J x CAST/EiJ
bp..... Base pair
C..... Cytosine
CAST..... CAST/EiJ
CB..... CAST/EiJ x C57BL/6J
ChIP..... Chromatin Immunoprecipitation
CpG..... Cytosine-guanine dinucleotide
ddH2O..... Double distilled water
DNA..... Deoxyribonucleic acid
DNMT..... DNA methyltransferase
dNTP..... Deoxynucleotide triphosphate
E..... Embryonic day
ER..... Endoplasmic reticulum
ERV..... Endogenous Retrovirus
ESC..... Embryonic stem cell
ETn..... Early transposon
g..... Gravitational force
G..... Guanine
GBE..... Genetic background effect
GV..... Germinal vesicle
H3K27ac..... Histone 3 Lysine 27 acetylation
H3K9me3..... Histone 3 Lysine 9 trimethylation

H4K20me3 ..... Histone 3 Lysine 20 trimethylation
HDAC ..... Histone deacetylase
HP1 ..... Heterochromatin protein 1
IAP..... Intracisternal A-Particle
ICM..... Inner cell mass
JAX ..... The Jackson Laboratory
KAP1 ..... KRAB-associated protein 1
kb ..... Kilo base
KRAB..... Krüppel-associated box
KZFP ..... KRAB zinc finger protein
LINE ..... Long interspersed element
LTR..... Long terminal repeat
MusD..... Type D endogenous provirus
NurD ..... Nucleosome remodeling and deacetylase
ORF..... Open reading frame
P ..... Postnatal day
PCR..... Polymerase chain reaction
PEV..... Position effect variegation
PGC ..... Primordial germ cell
piRNA ..... piwi-interacting RNA
RNA ..... Ribonucleic acid
RNAi ..... RNA interference
rpm ..... Revolutions per minute
RT-PCR ..... Reverse transcription PCR
SETDB1 ..... SET Domain Bifurcated Histone Lysine Methyltransferase 1
SINE ..... Short interspersed element
SNP..... Single nucleotide polymorphism
T ..... Thymine
TE..... Transposable elements
U ..... Uracil
UCSC..... University of California, Santa Cruz
VM-IAP..... Variably methylated IAP

Chapter 1

Introduction

While studying maize genetics in the 1950s, Barbara McClintock proposed that kernel colour variegation was driven by 'controlling elements' that could mobilise in the genome (McClintock 1950, 1951, 1953, 1956). Although the discovery of genetic transposition won her a Nobel Prize in 1983, her observations were met with scepticism at the time because they challenged the prevailing view of genes as static entities linearly arranged along chromosomes. Today, the prevalence of transposable elements (TEs) is undisputed. TEs make up more than 80% of the maize genome and account for approximately half of the mammalian genome (Schnable *et al.* 2009; Smit *et al.* 2015).

McClintock went on to describe 'changes of phase' displayed by TEs in maize (McClintock 1958). She reported oscillations between active and inactive gene expression phases that occurred in a developmentally regulated manner. It is now recognised that the phase changes she documented were epigenetic in nature and the product of fluctuating DNA methylation states. The study of TE regulation has thus become tightly linked to the field of epigenetics, which investigates heritable changes in gene expression that are not mediated by alterations to the underlying DNA sequence. This thesis lies at the intersection of these two fields and focuses on the epigenetic modification of TEs in a mammalian context.

1.1 The repeat genome

Repetitive DNA sequences are present in multiple copies in the genome. Current estimates suggest that repetitive DNA represents 45% and 52.5% of the mouse and human genomes, respectively (Smit *et al.* 2015). TEs account for the vast majority of mammalian repetitive DNA (**Figure 1.1**). As their name implies, TEs are capable of intracellular mobilisation, leading to multiple interspersed copies of the same element throughout the genome. Their abundance and diversity in eukaryotic genomes necessitates a hierarchical classification system (**Figure 1.2**; Wicker *et al.* 2007). Depending on their mechanism of transposition, TEs are classified as retrotransposons (type 1) or DNA transposons (type 2). Retrotransposons are transcribed into RNA intermediates which are reverse transcribed into DNA before insertion into a new target site. DNA transposons are excised by transposase enzymes and reinserted in a different location without passage through an RNA intermediate. TEs can also be classified as autonomous or non-autonomous: autonomous TEs carry all the necessary protein-coding sequences to mobilise independently whereas non-autonomous TEs rely on the machinery of another TE for successful transposition. These mechanisms of mobilisation lead to highly species-specific TE landscapes.

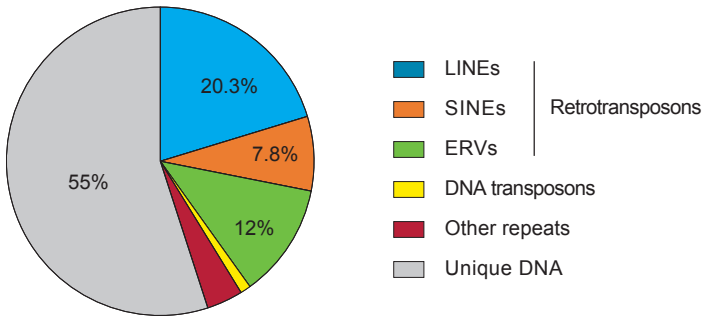


Figure 1.1: Content and distribution of the mouse repeat genome. Percentages were computed by RepeatMasker based on the GRCm38/mm10 genome assembly. Abbreviations: LINE, long interspersed element; SINE, short interspersed element; ERV, endogenous retrovirus.

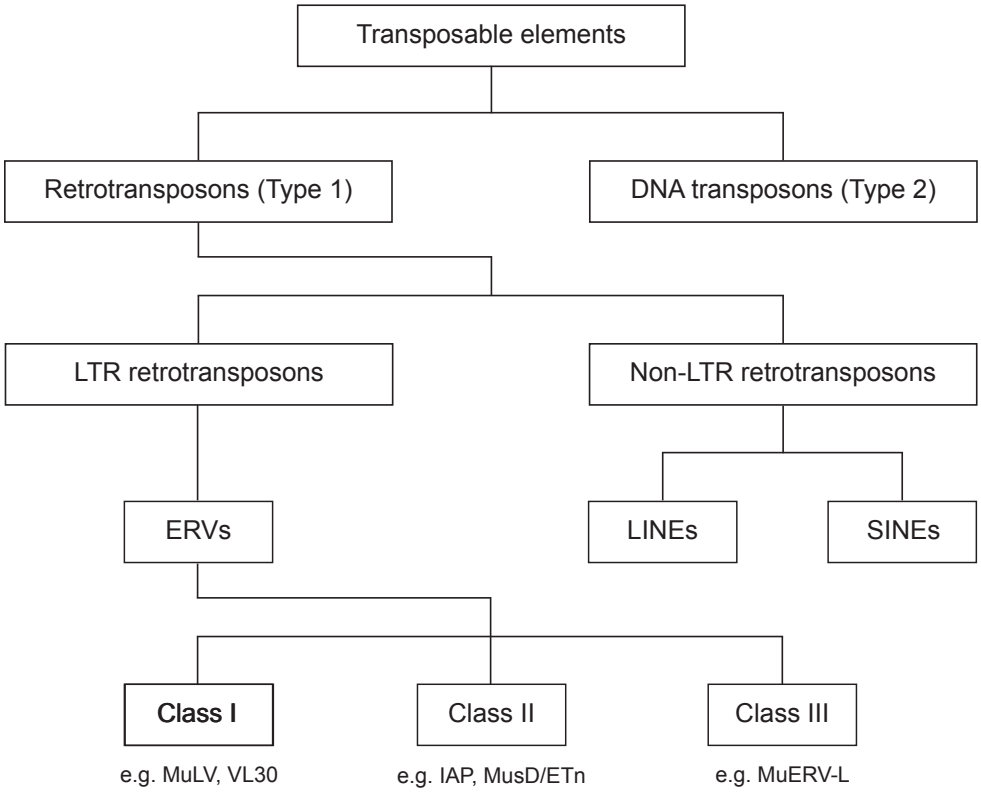


Figure 1.2: Hierarchical classification system for murine transposable elements (TEs). Abbreviations: LTR, long terminal repeat; ERV, endogenous retrovirus; LINE, long interspersed element; SINE, short interspersed element; MuLV, murine leukemia virus; VL30, virus-like 30; IAP, intracisternal A-particle; MusD, mus musculus type D; ETn, early transposon; MuERV-L, murine endogenous retrovirus L.

1.1.1 Retrotransposons

Retrotransposons make up 40% of the mouse genome, far outnumbering DNA transposons (Smit *et al.* 2015). They are categorised into long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons (**Figure 1.2**). LTR retrotransposons mostly consist of endogenous retroviruses (ERVs), discussed in more detail below. Non-LTR retrotransposons are made up of long or short interspersed elements (LINEs and SINES, respectively) (**Figure 1.1**). LINEs are the most prevalent retrotransposon in the mouse. Full-length LINEs are 6 to 7 kilobases (kb) in length and code for two open reading frames that catalyse autonomous retrotransposition, ORF1 and ORF2 (Boissinot & Sookdeo 2016). SINEs are non-autonomous elements ranging between 80 and 400 base pairs (bp)

in length. Their retrotransposition is carried out *in trans* by LINE-derived proteins (Dewannieux & Heidmann 2005).

ERVs make up 12% of the mouse genome and share sequence homology with retroviruses. In fact, any retroviral insertion that occurs in the germline and is subsequently inherited vertically as a provirus in host DNA is classified as an ERV. The internal sequence of full-length ERVs contains the *gag, pro, pol*, and *env* genes required for viral replication (**Figure 1.3**). The *gag* gene codes for a polyprotein necessary for viral particle assembly. The *pro* gene yields a protease enzyme involved in the maturation of retroviral proteins. The *pol* gene produces reverse transcriptase, RNase H, and integrase. The *env* gene codes for an envelope glycoprotein required for extracellular replication and infectivity. The ERV coding region is flanked by non-coding identical 5′ and 3′ LTR sequences. These contain three subregions: a regulatory region (R), a unique 5′ (U5) region, and a unique 3′ (U3) region (**Figure 1.3**). The tRNA primer binding site (PBS) between the 5′ LTR and the *gag* gene designates the site of reverse transcription initiation (Mager & Stoye 2015; Varmus 1982). ERV sequences contain a range of transcriptional regulation motifs, including transcription factor binding sites, polyadenylation signals, and splice acceptor sites (Maksakova *et al.* 2006). However, due to mutational decay, most ERVs no longer contain all of the proviral elements described above and are incapable of retrotransposition. Many exist as solo LTRs formed from homologous recombination between identical LTR sequences (Belshaw *et al.* 2007; Nellåker *et al.* 2012; Subramanian *et al.* 2011).

ERV elements are classified as Class I, Class II, or Class III ERVs based on their phylogenetic similarity to known exogenous retroviruses (**Figure 1.2**). Class II ERVs share homology with alpha- and betaretroviruses and have expanded considerably in the mouse; their occupancy is 10 times greater in mice compared to other mammals (Mager & Stoye 2015). Intracisternal A-particles (IAPs) and Early Transposon (ETn)/MusD elements are Class II ERVs and together account for the vast majority of reported TE-induced mutations in the mouse (Gagnier *et al.* 2019). This dissertation focuses on IAP elements.

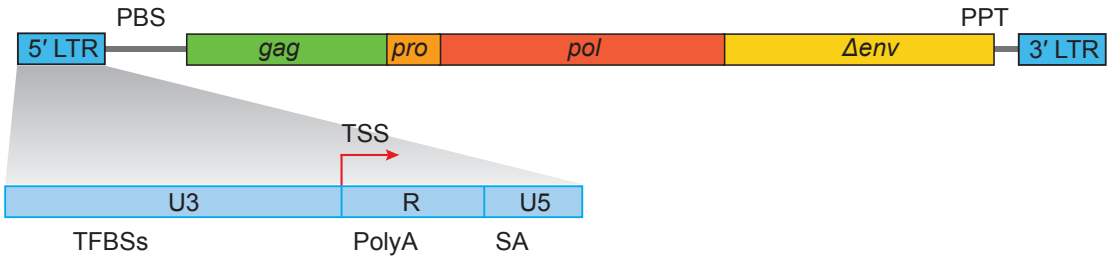


Figure 1.3: Structure of full-length endogenous retroviruses (ERVs). The order of the viral genes (*gag, pro, pol, env*) is conserved across ERVs. The *env* sequences in the vast majority of ERV elements are either truncated or absent (*Δenv*). Abbreviations: LTR, long terminal repeat; PBS, primer binding site; PPT, polypurine tract; TSS, transcription start site; U3, unique 3′ LTR region; R, regulatory LTR region; U5, unique 5′ LTR region; TFBSs, transcription factor binding sites; PolyA, polyadenylation signal; SA, splice acceptor site. Figure inspired from Jern & Coffin 2008.

1.1.2 Intracisternal A-particles (IAPs)

IAPs are evolutionarily young murine-specific ERVs present in approximately 12,000 copies in the C57BL/6J (B6) inbred mouse strain (Smit *et al.* 2015). They form virus-like particles that assemble in the cisternae of the endoplasmic reticulum (ER). These particles have been observed in tumour cells, thymus samples, and early embryos (Kuff & Lueders 1988). Unlike infectious ERVs capable of extracellular mobilisation, IAPs cannot exit the ER membrane because they lack a functional envelope glycoprotein (Perk & Dahlberg 1974). It is thought that the infectious retroviral progenitor of IAPs, itself still present in the mouse genome, gave rise to this class of ERVs via passive mutational loss of the *env* gene (Ribet *et al.* 2008). About 200 IAPs in the mouse genome still harbour *env*-like sequences, designated IAPEs, but only two of these are full-length elements with an intact *env* gene and identical LTRs (Reuss & Schaller 1991; Ribet *et al.* 2008). The loss of *env* was found to be correlated with a 30-fold increase in IAP proliferation, suggesting that the adoption of an intracellular replication mechanism facilitated their unusually rapid amplification in the mouse genome (Magiorkinis *et al.* 2012). In addition to the loss of *env*, most IAPs have a divergent domain in the *gag* gene causing the localisation of IAP particles to the ER instead of the cellular membrane (Ribet *et al.* 2008).

Although IAPs are theoretically capable of intracellular retrotransposition, approximately half of the IAPs in the mouse genome are solo LTRs and only a small minority have intact proviral genes. Historically, IAPs have been classified based on common structural variations. For example, the 1Δ1 IAP subclass carries a 1.9kb in-frame deletion between the *gag* and *pol* genes, resulting in a *gag-pol* fusion protein (Kuff & Lueders 1988). Given that both *gag* and *pol* are required for autonomous retrotransposition, the mobilisation of 1Δ1 elements relies on the transcription of *gag* and *pol* from other IAPs (Dewannieux *et al.* 2004). Counterintuitively, the majority of new IAP insertions are of the 1Δ1 subtype (Maksakova *et al.* 2006).

The most commonly used IAP classification system at present is the one established by the reference database Dfam and used by the Repeatmasker program for genome annotation. This system classifies IAP LTRs (e.g. IAPLTR1_Mm, IAPLTR2_Mm) and internal IAP portions (e.g. IAPEz-int, IAPEy-int) separately, and assigns them to a subclass based on sequence similarity. Unfortunately, the probabilistic model used for the classification process is obscure and the Dfam description for each subtype is uninformative.

Approximately a fifth of IAPs in the mouse genome have identical 5′ and 3′ LTR sequences (perfect-pair LTRs). This high LTR concordance and the mouse-specificity of IAP elements suggests that IAPs are evolutionarily young ERVs. The IAPLTR1 and IAPLTR1a subtypes are vastly over-represented among IAPs with perfect-pair LTRs, making them the youngest and most active subtypes (Qin *et al.* 2010). As a biproduct of their evolutionary youth, IAPs exhibit high degrees of insertional polymorphism across inbred mouse strains (Lueders *et al.* 1993; Nellåker *et al.* 2012; Zhang *et al.* 2008). A recent study catalogued all known TE-induced mutations in the mouse and determined their strain of origin (Gagnier *et al.* 2019). This meta-analysis revealed that 84% of IAP insertional mutations originated in the C3H mouse strain, suggesting IAPs are more active in C3H than in other strains. The cause for this is unknown but may involve altered transcriptional regulation, consistent with the fact that IAP expression levels in bone marrow, spleen, and thymus are higher in C3H than in B6 or STS/A mice (Ishihara *et al.* 2000).

1.2 Functional consequences of retrotransposition

As a consequence of their mobilising ability and inherent regulatory sequences, retrotransposons can disrupt genome function and contribute to disease progression. The effects of retrotransposition on gene expression are wide-ranging (**Figure 1.4**; Friedli & Trono 2015; Mager & Stoye 2015). Intragenic insertions can disrupt exons, create new isoforms via alternative splicing, and cause premature polyadenylation. The regulatory sequences embedded in LTRs can act as sense and antisense promoters of adjacent genes as well as provide new transcription factor binding sites (Gagnier *et al.* 2019; Rebollo *et al.* 2012). Retrotransposons can also influence gene expression in a *trans* capacity via enhancer activity or via transcription of regulatory RNAs (Thompson *et al.* 2016). Finally, the high sequence similarity between retrotransposons can result in aberrant recombination events such as duplications, deletions, and translocations (Mieczkowski *et al.* 2006).

Mutations caused by retrotransposons can have negative effects on host fitness. Somatic LINE-1 retrotransposition has been detected in a range of cancer tumours and 124 human genetic diseases have been linked to LINE-1-mediated insertions (Hancks & Kazazian 2016). Numerous repeat-induced deleterious mutations have also been documented in the mouse. Examples include a MusD insertion at the *Dac* locus that leads to limb malformation (Kano *et al.* 2007), an IAPEz insertion into the *Dnmt3C* gene that induces male sterility (Barau *et al.* 2016), and a LINE-1 insertion into the *Pde6C* gene that causes impaired vision (Chang *et al.* 2009). Host genome defence mechanisms have evolved to repress retrotransposon activity and mitigate these effects, as discussed in the next section.

Despite their potentially deleterious consequences, retrotransposons have also been co-opted over evolution by the host for adaptive purposes. For instance, the mammalian Syncytin cell fusion proteins, essential for placental development, are derived from the ERV *env* gene (Mi *et al.* 2000), and salivary amylase evolved in primates due to a primate-specific ERV insertion that drives amylase expression in salivary glands (Ting *et al.* 1992). In addition, murine retrotransposons have been reported to regulate host gene expression in oocytes and preimplantation embryos (Peaston *et al.* 2004), and more recent studies

suggest LTR-derived transcripts play a role in stem cell pluripotency (Fort *et al.* 2014; Macfarlan *et al.* 2012). The molecular domestication of retrotransposon sequences illustrates that, in principle, these elements can contribute to the evolution of biological pathways.

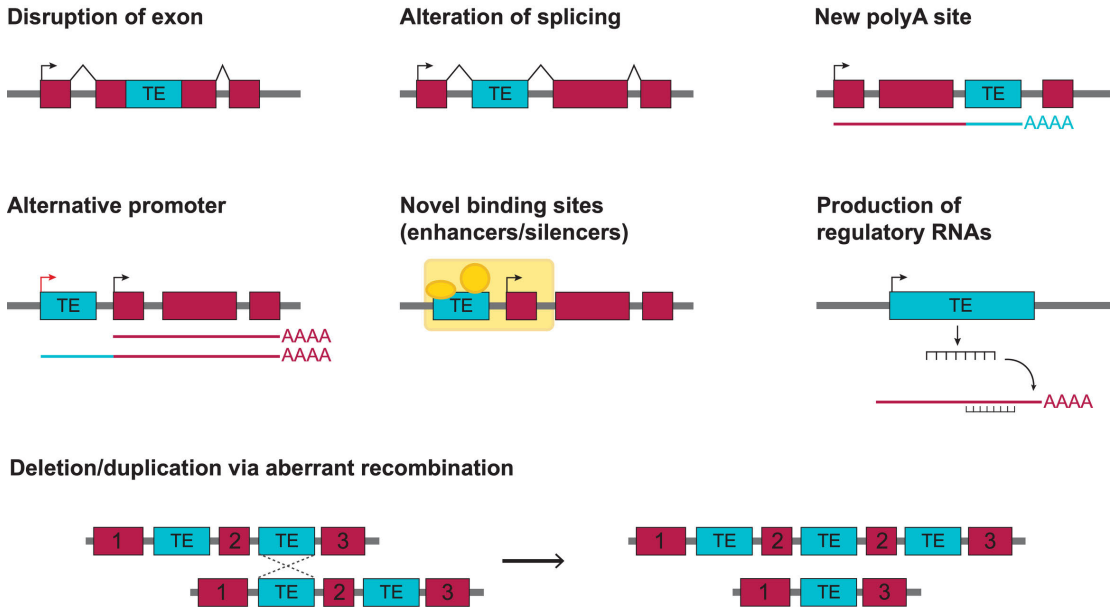


Figure 1.4: The possible effects of TEs on host genome function. TE insertions are shown in blue and host gene exons are shown in crimson. Figure adapted from Ecco *et al.* 2017.

The indication that TE insertions can have favourable effects on host fitness has been used as evidence against the selfish DNA model (Doolittle & Sapienza 1980; Orgel & Crick 1980). This model argues that the innate ability of TEs to mobilise and promote their own transmission fully accounts for their spread in the genome over evolution, rendering the improvement of host fitness non-essential for their survival. In reality, the TE co-option model and the selfish DNA model are not mutually exclusive. It is possible that a majority of TEs, which altogether account for half of the mammalian genome, are neutral in their effect on host fitness and have become fixed in the population via genetic drift. This does not preclude the hijacking of transposon machinery following chance

insertion events that improve host fitness, resulting in a symbiotic relationship between an individual TE and its host. Importantly, the inherent biochemical activity of TEs does not demonstrate a role in host genome function (Bourque *et al.* 2018). The development of sequencing and analytic technologies better optimised for the study of repetitive DNA will help determine the extent to which TE domestication has taken place in the mammalian genome. A thorough assessment of the functional significance of TEs must include large-scale studies that quantify purifying selection on TE sequences as well as more focused experiments investigating the phenotypic impact of losing specific elements.

1.3 Host defence mechanisms against retrotransposition

Mammals have evolved a range of mechanisms to mitigate the deleterious effects of retrotransposition. Many of these are epigenetic in nature in that they do not rely on changes to the retroelement DNA sequence to induce repression. Given that mutational decay has rendered most TEs incapable of autonomous mobilisation, the epigenetic mechanisms described below are particularly important for the repression of evolutionarily young elements.

DNA methylation, histone modification, non-coding RNAs, and DNA-binding proteins have all emerged as important and intertwined host defence mechanisms against retrotransposition. Of note, evolutionary theory indicates that retrotransposons are under simultaneous selection to (1) propagate in the host germline to ensure their survival in the next generation and (2) limit the extent of their mobilisation to avoid host death or infertility. This suggest that host-induced repression, particularly in the soma, may in certain cases be advantageous to retroelements and conducive to their successful amplification over evolution.

1.3.1 DNA methylation

DNA methylation involves the covalent addition of methyl (CH_3) groups to the fifth carbon of cytosine bases (5-methylcytodine, 5mC) and is mostly associated with transcriptional repression. The modification is found across

eukaryotes and is thought to have originated as a bacterial immune system (Bestor 1990; Zemach *et al.* 2010). A number of species have since lost DNA methylation, including most fungi and the model organisms *Drosophila melanogaster* and *Caenorhabditis elegans* (Zemach & Zilberman 2010). This may be related to the mutagenic effects of DNA methylation, as 5mC displays increased susceptibility to spontaneous conversion to thymidine (T) (Bird 1980; Wang *et al.* 1982). In mammals, DNA methylation occurs almost exclusively at cytosine-guanine dinucleotides (CpGs). 70 to 80% of CpGs in the mammalian genome are methylated in somatic tissues, with the notable exception of CpG islands (stretches of CpG rich DNA often found at promoters) (Bird 2002; Ehrlich *et al.* 1982).

The self-complementarity of CpGs allows methylation patterns to be faithfully copied onto newly synthesised DNA strands during cell division (Riggs 1975). This mitotic heritability is catalysed by the maintenance methyltransferase DNMT1 (DNA methyltransferase 1) (Bestor *et al.* 1988). DNMT1 is recruited to replication forks by the E3 ubiquitin-protein ligase UHRF1, which recognises hemimethylated CpGs (Bostick *et al.* 2007; Sharif *et al.* 2007). Because this mechanism relies on the presence of CpG methylation in the parental cell, DNMT1 is not capable of establishing new methylation patterns during early development and gametogenesis. This is the role of *de novo* DNA methyltransferases DNMT3A and DNMT3B (Okano *et al.* 1998, 1999). In both the male and female germlines, the activity of DNMT3A and DNMT3B relies on the binding of their catalytically inactive co-factor DNMT3L (Ooi *et al.* 2007). The importance of DNA methylation for normal development is highlighted by the severe phenotypic outcomes observed in DNA methyltransferase mutations in the mouse. *Dnmt1* mutants are embryonic lethal at embryonic day 9.5 (E9.5), *Dnmt3A* mutants exhibit postnatal lethality at 4 to 8 weeks of age, and *Dnmt3B* mutants are embryonic lethal at E14.5 (Li *et al.* 1992; Okano *et al.* 1999). Deletion of *Dnmt3L* causes male sterility and heterozygous embryos produced from *Dnmt3L* mutant females die at E9.5 (Bourc'his *et al.* 2001).

1.3.1.1 Biological functions of DNA methylation

It has been proposed that DNA methylation originally evolved as a host defence mechanism against TEs that was subsequently co-opted as a repressive mark in other biological contexts (Yoder *et al.* 1997; Zemach *et al.* 2010). Indeed, TEs are highly methylated in mammalian genomes. Treating mice with 5-azacytidine (a toxic DNA methylation inhibitor) induces retroviral transcription in postnatal somatic tissues (Jaenisch *et al.* 1985), though the extent of this on a per element basis is not known. *Dnmt1*-deficient embryos show a striking increase in IAP transcription (Walsh *et al.* 1998) and loss of *Dnmt3L* results in reactivation of both LTR and non-LTR retrotransposons in the male germline (Bourc'his & Bestor 2004). Lack of *Dnmt3L* also leads to abnormal chromosome pairing, suggesting DNA methylation may mask otherwise homologous repetitive sequences to prevent erroneous pairing during cell division. In addition, mutations in the recently identified rodent-specific *de novo* DNA methyltransferase gene *Dnmt3C* cause an upregulation of evolutionarily young retrotransposons in the testis (Barau *et al.* 2016; Jain *et al.* 2017). Finally, the mutagenic properties of DNA methylation increase the rate of C-to-T transitions in repeat elements, thereby accelerating the disruption of their retrotransposition potential.

DNA methylation regulates other biological functions in mammals in addition to the mitotically heritable silencing of TEs. These include genomic imprinting and X chromosome inactivation. Genomic imprinting is the process by which germline-specified differential methylation results in monoallelic expression of a gene based on its parental origin (Ferguson-Smith 2011). Mammalian X chromosome inactivation is the gene dosage compensation mechanism leading to stable transcriptional repression of one of the female X chromosomes (Xi), and is maintained by DNA methylation which is secondary to other repressive mechanisms (Chaligné & Heard 2014). However, X chromosome and imprinted genes aside, there is little evidence for a widespread and instructive role for DNA methylation in gene expression regulation. Many studies have shown correlations between gene promoter demethylation and transcriptional activation, but few (if any) have shown that this association is causative (Bestor *et al.* 2015; Schübeler 2015). Interestingly, DNA methylation is

enriched at gene bodies, especially exons (Lister *et al.* 2009). Gene body methylation represents the most evolutionarily conserved methylation patterning across eukaryotes but does not appear to be essential (Feng *et al.* 2010; Zemach *et al.* 2010). Additional work is required to determine its functional significance, with current hypotheses suggesting roles in the masking of cryptic promoters and in the regulation of alternative splicing via the methylation sensitive DNA-binding protein CTCF (CCCTC binding factor) (Gelfman *et al.* 2013; Neri *et al.* 2017; Shukla *et al.* 2011).

DNA methylation has emerged as a powerful biomarker for disrupted cellular states. Global hypomethylation and CpG island hypermethylation were observed in tumour samples decades ago (De Bustros *et al.* 1988; Feinberg & Vogelstein 1983; Toyota *et al.* 1999). Abnormal DNA methylation patterns have since been reported in a range of cancer types and have resulted in the therapeutic use of DNA methylation inhibitors (reviewed in Yang *et al.* 2010). Furthermore, the hope is that methylation biomarkers will prove useful for early detection and diagnosis in the clinic (reviewed in Mikeska & Craig 2014). DNA methylation has also been shown to change predictably over the human and mouse lifespans at specific CpG sites. This has led to the development of 'DNA methylation clocks' capable of predicting age based on DNA methylation data alone (Bocklandt *et al.* 2011; Hannum *et al.* 2013; Horvath 2013; Petkovich *et al.* 2017; Stubbs *et al.* 2017; Wang *et al.* 2017; Weidner *et al.* 2014). These clocks may have applications in the detection of accelerated ageing and in the identification of factors that influence the ageing process.

1.3.1.2 Genome-wide reprogramming of the methylome

The mammalian methylome undergoes two rounds of genome-wide reprogramming, once during preimplantation development and again during germline specification (**Figure 1.5**; Monk *et al.* 1987; Sanford *et al.* 1987). Immediately after fertilisation, the highly condensed paternal genome is stripped of protamines and reassembled into nucleosomes using oocyte-derived histones (McLay & Clarke 2003). Soon after, germline-specific methylation patterns are lost from both the paternal and maternal genomes via global demethylation (Santos *et al.* 2002). The paternal genome is subject to both active

and passive demethylation. Active demethylation occurs via oxidation of 5mC to 5-hydroxymethylcytosine (5hmC) by the methylcytosine dioxygenase TET3 (Gu *et al.* 2011). Passive demethylation involves the replication-dependent dilution of DNA methylation in the absence of DNMT1, which is excluded from the nucleus during early embryonic cleavages (Carlson *et al.* 1992). The maternal genome is mostly reprogrammed via passive demethylation, although recent evidence suggests that TET3-mediated oxidation also occurs on the maternal genome to a lesser extent (Guo *et al.* 2014; Shen *et al.* 2014). In both mice and humans, the lowest level of global DNA methylation is reached at the blastocyst stage (E3.5 in mice). After implantation, cell lineage-specific methylation patterns are established by *de novo* methyltransferases.

The second wave of global demethylation occurs in primordial germ cells (PGCs) after their emergence from the epiblast (E7.25 in mice). In both male and female PGCs, an initial phase of global passive demethylation is followed by TET1- and TET2-mediated locus-specific active demethylation, resulting in a largely unmethylated genome by E13.5 (Hackett *et al.* 2013; Seisenberger *et al.* 2012; Vincent *et al.* 2013). Subsequent *de novo* DNA methylation occurs at distinct stages and to different degrees in the male and female germlines, resulting in sex-specific germline methylation profiles, including differentially methylated imprinting control regions (ICRs). Male germ cells become highly methylated before birth and prior to reaching full maturation. Female germ cells initiate *de novo* methylation after birth during ovulation (Seisenberger *et al.* 2013). Final global methylation levels are significantly lower in oocytes (~50%) than in sperm (~80%) (**Figure 1.5**; Wang *et al.* 2014).

This two-wave model of mammalian epigenetic reprogramming is not applicable to the entire murine genome: approximately 20% and 10% of CpGs are methylated at the blastocyst stage and in E13.5 PGCs, respectively (Wang *et al.* 2014). This suggests that some sequences are refractory to methylation reprogramming. The best-described example of resistance to demethylation occurs at genomic imprints. Differential methylation established at ICRs during germline specification is retained during preimplantation reprogramming, leading to parent-of-origin-specific monoallelic gene expression in the developing embryo (Ferguson-Smith 2011). This protection of imprinted regions is mediated by the Krüppel-associated box (KRAB)-containing zinc finger

proteins ZFP57 and ZFP445 (Li *et al.* 2008; Strogantsev *et al.* 2015; Takahashi *et al.* 2019). Some evolutionarily young retrotransposons also resist reprogramming, particularly IAPs. IAPs can remain methylated (though not fully) during both preimplantation development and germline specification (Lane *et al.* 2003; Seisenberger *et al.* 2012). This may reflect the high cost of derepressing retrotransposition-competent elements. It has been proposed that resistance to epigenetic reprogramming may underlie instances of non-genetic inheritance of phenotypic traits induced in a previous generation.

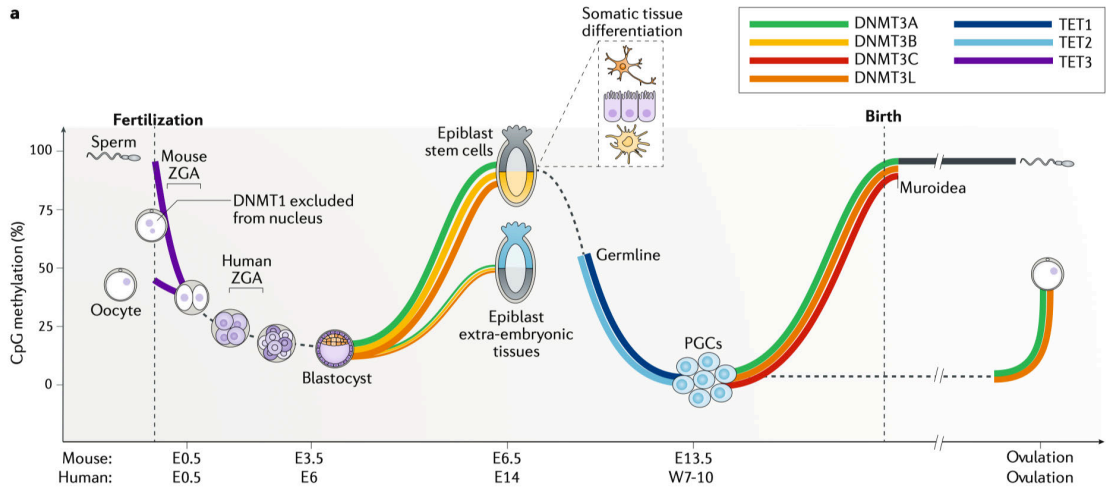


Figure 1.5: Genome-wide epigenetic reprogramming in mammals. The mammalian methylome undergoes two waves of erasure during development. The first wave of demethylation begins immediately after fertilisation and continues until the blastocyst stage, followed by the establishment of cell-type specific DNA methylation patterns. The second wave occurs in primordial germ cells (PGCs) and allows for the acquisition of sex- and germ cell-specific methylation signatures. Figure from Greenberg & Bourc'his 2019.

1.3.2 Histone modification

Genome organisation and compaction relies on the wrapping of DNA around histone proteins to form nucleosomes. DNA and its associated histones are

referred to as chromatin. The residues on the N-terminal tails of histones are post-translationally modified by a range of chemical modifications, including methylation and acetylation. The distribution of such marks along the genome has important consequences for chromatin structure and genome function, although the specific role of each modification, the total number of which is ever-expanding, is not fully understood (Zhao & Garcia 2015). Widespread chromatin reorganisation occurs during the two waves of genome-wide demethylation described in the previous section, but insight into these processes is limited due to the technical difficulties associated with studying chromatin organisation at these developmental stages when cell number is a restricting factor.

Only a few histone modifications have been reported to associate with retrotransposons. In mouse embryonic stem cells (ESCs), ERVs are enriched for histone H3 lysine 9 trimethylation (H3K9me3), histone H4 lysine 20 trimethylation (H4K20me3), and histone 3 lysine 64 trimethylation (H3K64me3), all of which are marks for transcriptional repression (Daujat *et al.* 2009; Martens *et al.* 2005; Mikkelsen *et al.* 2007). Knockout experiments in mouse ESCs showed that the H3K9 methyltransferase SETDB1 (SET domain bifurcated 1, also known as ESET) is required for ERV silencing but that the H4K20 methyltransferases SUV420H1 and SUV420H2 are not, which suggests that H3K9me3 plays a primary role in this process (Matsui *et al.* 2010). Unlike *Setdb1* deficiency, loss of *Dnmt1* does not result in ERV upregulation in mouse ESCs (Hutnick *et al.* 2010; Karimi *et al.* 2011). This indicates that ERV repression in ESCs is mediated by an H3K9me3-dependent and methylation-independent mechanism that may be reflective of ERV regulation in the unmethylated preimplantation embryo. Simultaneous deletion of both *Setdb1* and *Dnmt1* in ESCs does not lead to greater ERV activation compared to deletion of *Setdb1* alone, with the exception for the IAPEz subclass of IAP elements which show a cumulative effect in double knockout cells (Karimi *et al.* 2011). These IAPs may represent the elements that are resistant to global demethylation during early development. Of note, transient retroelement transcription has been reported in one- and two-cell embryos, suggesting that at least a subset of elements are epigenetically derepressed immediately after fertilisation (Kigami *et al.* 2003; Peaston *et al.* 2004). In an attempt to mimic early embryo dynamics, Walter and colleagues induced DNA demethylation in mouse ESCs by transferring the cells from

serum-based to 2i+vitC (MEK and GSK3 inhibition with vitamin C) medium (Walter *et al.* 2016). Following the transfer, a progressive accumulation of histone 3 lysine 27 trimethylation (H3K27me3) was observed at full-length elements of various ERV families, but not at IAPEz elements. In addition, mouse ESCs lacking functional Polycomb-repressive complexes 1 and 2 (PRC1 and PRC2) show upregulation of IAP and Murine Leukemia Virus (MuLV) elements (Leeb *et al.* 2010). Given that PRC2 catalyses H3K27 methylation, these studies suggest that in certain conditions H3K27me3 may play a role in ERV repression.

The active removal of histone modifications that associate with transcriptional activation has also been implicated in ERV repression. Chemical inhibition of histone deacetylases (HDACs) with trichostatin A (TSA) in mouse fibroblasts results in an increase in H3 acetylation at SINE and virus-like 30 (VL30) LTR elements as well as transcriptional upregulation of VL30 and IAP LTR retrotransposons (Brunmeir *et al.* 2010).

1.3.3 Pathways mediating the epigenetic repression of retrotransposons

The mechanisms that recruit repressive epigenetic marks to retrotransposons following genome-wide reprogramming are currently the topic of intense investigation. RNA-mediated mechanisms and KRAB zinc finger proteins (KZFPs) have both emerged as important mediators of sequence-specific epigenetic repression of retrotransposons.

1.3.3.1 RNA-mediated mechanisms

The piRNA (piwi-interacting RNA) pathway was first discovered in *Drosophila* and has since been identified in species across the animal kingdom. In brief, piRNA biogenesis involves the export of retrotransposon-derived single-stranded piRNA precursors from the nucleus into the cytoplasm, where they are cleaved into 24 to 35 nucleotide fragments and loaded onto P-element-induced wimpy testis (piwi) proteins. Secondary processing via the ping pong cycle results in successive rounds of piRNA amplification. These piRNA-protein complexes are capable of re-entering the nucleus to direct the epigenetic

repression of TEs. Thus, piRNAs mediate both pre- and posttranscriptional TE regulation (reviewed in Czech & Hannon 2016).

Studies over the past two decade have shown that the mammalian piRNA pathway plays a crucial role in silencing young retrotransposons in the male germline. Transposon-derived piRNAs are readily detected in testes and all three of the non-redundant piwi genes encoded in the mouse genome (*Mili*, *Miwi*, and *Miwi2*) are required for the successful progression of spermatogenesis (Aravin *et al.* 2007; Carmell *et al.* 2007; Deng & Lin 2002; Kuramochi-Miyagawa *et al.* 2004). Loss of either *Mili* or *Miwi2* in male germ cells results in a 5- to 10-fold increase in LINE-1 and IAP expression (Aravin *et al.* 2007; Carmell *et al.* 2007). This is accompanied by a decrease in DNA methylation at these elements, suggesting piRNAs may drive the sequence-specific targeting of retrotransposons for *de novo* DNA methylation in PGCs (Aravin *et al.* 2007; Kuramochi-Miyagawa *et al.* 2008). Indeed, piRNA-bound Miwi2 localises to the nucleus during the *de novo* methylation stage of spermatogenesis (Aravin *et al.* 2008). Loss of *Dnmt3L*, itself essential for the methylation of retrotransposons in the male germline by DNMT3A, does not impede piRNA expression (Aravin *et al.* 2008). This supports a role for the piRNA pathway upstream of DNMT3L-dependent *de novo* DNA methylation and provides a mechanism for the sequence-specificity of the process. However, more recent work suggests that the vast majority of LINE and LTR elements undergo *de novo* DNA methylation in the male germline independently of the piRNA pathway, with loss of *Mili* and *Miwi2* only interfering with *de novo* DNA methylation of a small subset of evolutionarily young retroelements (Molaro *et al.* 2014). The piRNA pathway has also been implicated in the deposition of H3K9me3 over LINEs in the male germline, but this was restricted to full-length active LINE-1s (Pezic *et al.* 2014).

In other model organisms such as *Drosophila,* piwi proteins are required for both spermatogenesis and oogenesis (Rojas-Ríos & Simonelig 2018). In contrast, mutations in piwi genes in the mouse result in male sterility but have no effect on female fertility (Carmell *et al.* 2007; Deng & Lin 2002; Kuramochi-Miyagawa *et al.* 2004). Although transposon-derived piRNAs have been detected in mouse oocytes and *Mili* knockout female germ cells display an increase in IAP transcription, both of these effects are less pronounced compared to those observed in male germ cells (Tam *et al.* 2008; Watanabe *et al.* 2008). Interestingly,

a recent study on human, macaque, and bovine ovaries showed much higher levels of piwi and piRNA expression compared to those reported in mice (Roovers *et al.* 2015). It has been suggested that the piRNA and RNA interference (RNAi) pathways act semi-redundantly in the mouse oocyte. RNAi involves the Dicer-mediated cleavage of double stranded RNA into small interfering RNAs (siRNAs), which subsequently direct the degradation of complementary mRNA sequences. Transposon-derived siRNA populations are detected in mouse oocytes and *Dicer* knockout oocytes exhibit upregulation of RLTR10 ERV elements (Tam *et al.* 2008; Watanabe *et al.* 2008). RNAi may also be involved in retrotransposon repression in the early mouse embryo because *Dicer* knockdown results in an increase in IAP and MuERV-L transcripts at the 8-cell stage (Svoboda *et al.* 2004). However, additional experiments are required to more clearly elucidate the role of RNA-mediated mechanisms in retrotransposon silencing in the female germline and preimplantation embryo.

1.3.3.2 KRAB zinc finger proteins (KZFPs)

KZFPs are highly specific DNA binding proteins representing the largest transcription factor family in higher vertebrates. Their array of C_2H_2 zinc fingers recognizes nucleic acid motifs and their KRAB domain recruits KAP1 (also known as TRIM28 and TIF1β), which in turn triggers heterochromatin formation via H3K9 methylation and DNA methylation. KAP1-interacting factors include Heterochromatin protein 1 (HP1), the histone deacetylating NuRD/HDAC complex, the H3K9 methyltransferase SETDB1, and DNA methyltransferases (**Figure 1.6**; reviewed Ecco *et al.* 2017). KAP1 depletion in human and mouse ESCs leads to loss of H3K9me3 and upregulation of ERVs (Rowe *et al.* 2010; Turelli *et al.* 2014). A genome-wide study reported that a majority of human KZFPs bind retrotransposons (Imbeault *et al.* 2017). This occurs in a combinatorial fashion: multiple KZFPs are able to recognise a single retrotransposon subfamily and a single KZFP is able to bind multiple retrotransposon subfamilies. A handful of studies have characterised the role of individual human or mouse KZFPs in retrotransposon regulation. In humans, SVA and LINE-1 elements are repressed by the primate-specific ZNF91 and ZNF93 proteins, respectively (Jacobs *et al.* 2014). In the mouse, ZFP809 primarily

targets ERV1 elements, ZFP819 silences IAP, SINE, and LINE retrotransposons, the paralogues ZFP932 and Gm15446 regulate ERVK elements, and the ZFP Gm6871 represses LINE-1 elements (Castro-Diaz *et al.* 2014; Ecco *et al.* 2016; Tan *et al.* 2013; Wolf & Goff 2009; Wolf *et al.* 2015). More recently, epigenetic repression of RMER19B retrotransposons was attributed to ZFP708 (Seah *et al.* 2019). Together, these studies highlight the conservation of TE repression as a core KZFP function.

KZFPs arose in the common ancestor of coelacanths and tetrapods and are present in the hundreds in most species analysed to date, except for birds. Despite their comparable number across species, KZFPs are highly species- and class-specific, indicative of rapid evolution and amplification (Imbeault *et al.* 2017). In vertebrate genomes, LTR retrotransposon abundance is correlated to the number of host *Zfp* genes (Thomas & Schneider 2011). This supports an arms race model by which retrotransposons mutate to evade KZFP repression, which in turn prompts the selection of novel KZFPs that can bind to the 'escapee' elements. The selective pressures resulting from this competition are thought to drive their co-evolution and increase genome complexity. Specific instances supporting this model have been identified. For example, human ZNF93 targets L1PA LINE elements but is unable to bind newer L1PA subfamilies due to deletion of the ZNF93 binding site (Jacobs *et al.* 2014). A similar relationship with L1PA elements was later observed for ZNF141, ZNF649, and ZNF765, where zinc-finger mutations appeared concomitant with new L1PA subfamilies (Imbeault *et al.* 2017). However, the latter study found widespread KZFP binding of degraded elements that have lost the ability to mobilise in the genome and reported examples where the invasion of TEs in the human genome occurred after, not before, the emergence of KZFPs capable of binding to them. The same study identified extensive tissue-specific expression of KZFPs. Based on these findings, the authors propose that the evolutionary arms race model is too simplistic and put forward a domestication model whereby KZFPs aid in the domestication of TEs and contribute to species-specific gene regulation (Ecco *et al.* 2017).

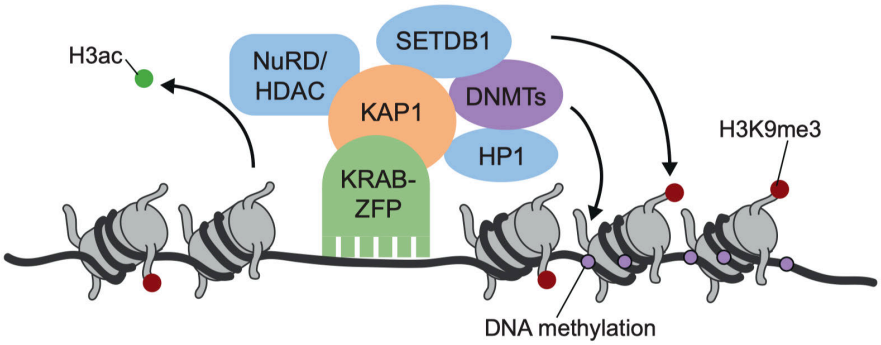


Figure 1.6: Heterochromatin formation by the KZFP-KAP1 complex. Following the recognition of a DNA motif by the KZFP zinc finger array, the KRAB domain recruits KAP1, which serves as a scaffold for the sequential recruitment of SETDB1, the NuRD/HDAC complex, HP1, and DNMTs. This results in transcriptional repression of the target site via H3 deacetylation, H3 K9 methylation, and DNA methylation. Figure from Ecco *et al.* 2017.

While the targets of most human KZFPs are TEs, the biological functions of KZFPs extend beyond the repression of parasitic DNA. As previously mentioned, ZFP57 and ZNF445/ZFP445 protect ICRs from demethylation during preimplantation epigenetic reprogramming (Li *et al.* 2008; Strogantsev *et al.* 2015; Takahashi *et al.* 2019). Other non-TE related KZFP functions include roles in metabolism (e.g. ZFP69 and ZN224) and cellular differentiation (e.g. ZFP689, ZFP13, ZNF268) (Barde *et al.* 2013; Chung *et al.* 2015; Lupo *et al.* 2011; Zeng *et al.* 2012). The next decade is likely to be a prolific period for studies focused on both specific KZFPs and their corresponding targets as well as additional *en masse* studies within and across species that further characterise the unique capabilities and evolutionary origins of this family of transcription factors.

1.3.3.3 The HUSH complex

A recent forward genetic screen for regulators of transgene position-effect variegation (PEV) in human cells identified the human silencing hub (HUSH) complex as a novel epigenetic repressor of retroviruses (Tchasovnikarova *et al.*

2015). The HUSH complex is conserved from fish to humans and is made up of three subunits: TASOR (transgenic activation suppressor), MPP8 (M-phase phosphoprotein 8), and periphilin. The complex localises to regions of the genome enriched for H3K9me3 and promotes the spreading of this mark via the recruitment of the H3K9 methyltransferase SETDB1 (Tchasovnikarova *et al.* 2015). HUSH-mediated heterochromatinisation relies on the recruitment and ATPase activity of MORC2 (microchidia CW-type zinc-finger 2) (Tchasovnikarova *et al.* 2017). A study on mouse naïve pluripotent stem cells showed that HUSH components cooperate with KAP1 to repress a subset of young LINE-1 retrotransposons (Robbez-Masson *et al.* 2018). Unlike KAP1, which relies on KZFP-mediated recognition of target sequences to recruit SETDB1, the HUSH complex recruits SETDB1 to heterochromatic loci in a sequence-agnostic manner.

1.4 Metastable epialleles

Mammalian TEs are rarely unmethylated in fully differentiated cell types. Rarer still are TEs that exhibit variable DNA methylation levels in the absence of genetic or environmental variation. Metastable epialleles fall into this unusual category of genomic loci. They have been regarded as examples of transgenerational epigenetic inheritance and potential biomarkers of environmental compromise.

1.4.1 Classic metastable epialleles: *A^{vy}* and *$Axin^{Fu}$*

In order for the definition of metastable epialleles to become clear, let us first consider two classic examples in the mouse: the *Agouti viable yellow* (*A^{vy}*) and *Axin fused* (*$Axin^{Fu}$*) alleles. Both exhibit abnormal gene expression due to IAP insertions (Duhl *et al.* 1994; Vasicek *et al.* 1997). The *Agouti viable yellow* (*A^{vy}*) allele arose from the spontaneous insertion of an IAP into pseudoexon1A (PS1A) of the *Agouti* coat colour gene locus (**Figure 1.7A** and **B**; Dickies 1962; Duhl *et al.* 1994; Waterland & Jirtle 2003). PS1A is located approximately 100 kb upstream of the *Agouti* coding exons (**Figure 1.7B**). Wild type *Agouti* is normally expressed transiently from a hair cycle-specific promoter, producing a paracrine signalling peptide that yields a yellow band on a black hair (Miller *et al.* 1993). This tightly controlled expression pattern is responsible for brown wild-type 'agouti' fur. The Agouti peptide can also interfere with metabolic pathways and has been linked to obesity, glucose intolerance, and tumourigenesis (Miltenberger *et al.* 1997; Yen *et al.* 1994). In *A^{vy}* mice, a cryptic promoter in the IAP LTR drives constitutive ectopic *Agouti* expression (**Figure 1.7B**; Duhl *et al.* 1994). This produces a mouse with a completely yellow coat as well as adult-onset obesity and diabetes (Yen *et al.* 1994). The *A^{vy}* allele is part of a series of dominant *Agouti* alleles brought about by IAP insertions, suggesting the *Agouti* gene locus may be particularly prone to insertional mutagenesis. Some of these alleles, such as *A^{iapy}* and *A^{hvy}*, cause similar phenotypes to *A^{vy}* (Argeson *et al.* 1996; Michaud *et al.* 1994). These additional dominant *Agouti* alleles have not been studied as extensively as *A^{vy}* and will not be discussed further.

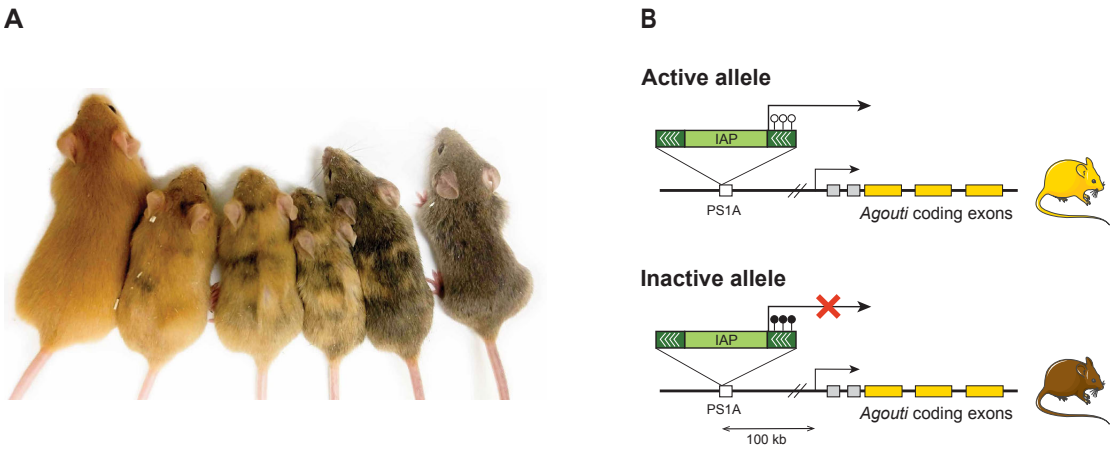


Figure 1.7: The *Agouti viable yellow* (A^{vy}) locus. A. Genetically identical A^{vy} mice display a range of coat colours from yellow to pseudoagouti, including varying levels of mottling. The mice shown are on a B6 background and are not age-matched siblings. **B.** The A^{vy} allele is characterised by the presence of a contra-oriented IAP insertion (directionality shown with white arrows) in pseudoexon1A (PS1A) 100kb upstream of the *Agouti* coding exons. The IAP is variably methylated between individuals and drives constitutive *Agouti* expression when unmethylated, leading to yellow fur. Full methylation results in pseudoagouti fur and partial methylation gives rise to mottling. Figure from Bertozzi & Ferguson-Smith 2019.

The *Axin fused* ($Axin^{Fu}$) allele resulted from an IAP insertion in the sixth intron of the *Axin* gene (Reed 1937; Vasicek *et al.* 1997). The Axin protein is involved in the regulation of embryonic axis formation by inhibiting Wnt signaling (Zeng *et al.* 1997). The intragenic position of the IAP causes aberrant transcription of *Axin* downstream exons, producing an atypical protein. The resulting truncated Axin interferes with axial patterning and results in the development of a distinctive kinked tail (Rakyan *et al.* 2003; Vasicek *et al.* 1997). Of note, while the $Axin^{Fu}$-associated IAP is intragenic, the one associated with the A^{vy} locus is 100kb upstream of the affected exons, supporting the previously mentioned concept that retrotransposon-mediated gene regulation can occur both locally and from a distance.

Unlike most mutations, A^{vy} and $Axin^{Fu}$ mice show variable penetrance and expressivity despite genetic homogeneity. The coat colour of individual inbred A^{vy} mice ranges from completely yellow to seemingly wild-type agouti (termed *pseudoagouti*), including varying intermediate degrees of mottled patterns (**Figure 1.7A**; Morgan *et al.* 1999). Likewise, tail morphologies in inbred $Axin^{Fu}$

mice span from straight to severely kinked (Rakyan *et al.* 2003). In both cases, the mechanism underlying the continuous phenotypic spectra is epigenetic in origin, with DNA methylation at the IAPs inversely correlated with expressivity. Hypomethylation at the IAP LTR thus corresponds to increased allelic expression, and vice versa (Morgan *et al.* 1999; Rakyan *et al.* 2003). Remarkably, the full span of phenotypes can be observed within a single litter regardless of parental phenotype, illustrating the instability of their epigenetic state after passage through the germline (Dickies 1962; Reed 1937).

The term *metastable epiallele* should make more sense now. The word *metastable* was first used in this context by plant biologists to describe alleles whose epigenetic state is capable of switching between active and repressed states from one generation to another (Hollick *et al.* 1995; Miura *et al.* 2009; Styles & Brink 1966). Adapted for mammals by Emma Whitelaw and colleagues in 2002, the term *metastable epiallele* is intended to highlight (1) the epigenetic basis for the phenotypes associated with these alleles and (2) the apparent stochasticity of their epigenetic state (Rakyan *et al.* 2002). In practice, this has translated to methylation variation between genetically identical individuals and, importantly, consistency in methylation levels within a single individual. While the methylation level of the IAPs associated with A^{vy} and $Axin^{Fu}$ varies across different mice, it is constant across different tissues of a single mouse, suggesting the methylation state is established early in development before tissue differentiation and maintained mitotically thereafter (Waterland *et al.* 2006; Waterland & Jirtle 2003). This differentiates metastable epialleles from differentially methylated regions (DMRs), a broader term used to designate differential but invariant methylation between biological samples, which could be cells, tissues, or individuals depending on the context.

Methylation consistency across tissues at the A^{vy} locus may seem at odds with the variegated patches of yellow and agouti fur observed on mottled mice. One might have expected pelts of intermediately methylated individuals to display a consistent intermediate pigmentation. DNA methylation, however, is dichotomous. A single CpG site is either methylated or unmethylated. Hence, the evident inter-individual range of methylation results from different proportions of methylated alleles in a cell population. Following from that, the proportion of methylated cells at A^{vy} is likely determined in a probabilistic

fashion before germ layer differentiation and propagated mitotically throughout the body as it develops (Waterland & Jirtle 2003; Yen *et al.* 1994). This would result in an approximately equal methylation percentage across tissues but would allow for local patches of cells to be different depending on the methylation state of their clonal origin. This is reminiscent of the black and orange mosaic fur pigmentation observed in tortoiseshell cats due to X-chromosome inactivation (Lyon 1961). Given that cellular development and proliferation differ between cell types, it is more accurate to think of metastable epialleles as loci that display a substantial correlation in methylation between tissues rather than identical intra-tissue methylation.

1.4.2 Assessing the prevalence of metastable epialleles genome-wide

1.4.2.1 Murine metastable epialleles

As discussed above, retroelement insertions play a key role in the unique behaviour of the A^{vy} and $Axin^{Fu}$ loci. Considering almost half of the mouse genome is made up of repetitive elements, it is perhaps unsurprising that other metastable epialleles have been identified. While A^{vy} and $Axin^{Fu}$ were discovered decades ago due to their striking visual phenotypes (Dickies 1962; Reed 1937), the identification of additional candidates has relied on using the genetic and epigenetic features of these classic loci to develop genome-wide screens and search algorithms. The third metastable epiallele to be identified was discovered by inspecting B6 cDNA databases in the hopes of finding transcripts containing IAP LTR sequences (Druker *et al.* 2004). One such sequence contained a contra-oriented IAP element in the sixth intron of the *Cdk5rap1* gene. Much like the A^{vy} and $Axin^{Fu}$ loci, this new candidate (named $Cabp^{IAP}$) shows an inverse correlation between IAP LTR methylation and expression of the aberrant transcript initiated from the 5′ LTR. The inter-individual methylation range of $Cabp^{IAP}$ (~20%) is much narrower than those observed at A^{vy} and $Axin^{Fu}$, and no identifiable phenotype is associated with its epigenetic variability (Druker *et al.* 2004).

The advent of high-throughput sequencing in the past decade has enabled larger-scale screens for metastable epialleles. The first attempt searched the

mouse genome using genome-wide expression microarray data. Transcripts exhibiting wide-ranging inter-individual variation and low-ranging inter-tissue variation were selected as candidates in an attempt to capture the expression pattern observed for the *Agouti* gene in A^{vy} mice (Weinhouse *et al.* 2011). Only two loci (*Dnajb1* and *Glcci1*) were analysed in depth and though they showed inter-individual methylation differences, neither exhibited methylation-associated expression.

The second attempt screened for IAP insertions with promoter activity. The study identified retrotransposons near mRNA promoters associated with H3K4me3, an activating histone modification (Ekram *et al.* 2012). This enriched for active IAP LTR promoters and resulted in a set of 143 candidate regions, from which 13 were selected for experimental validation. Only three of these were found to exhibit significant methylation variation between individuals, with ranges of approximately 50% (Ekram *et al.* 2012). Also focusing on repeat elements, Faulk and colleagues recognized that the IAPs associated with A^{vy} and $Cabp^{IAP}$ both belong to the IAPLTR1_Mm subclass. They showed that IAPLTR1_Mm elements cluster into three clades, with the largest one containing the most conserved elements. A^{vy} and $Cabp^{IAP}$ segregated together in a separate smaller clade. Based on a limited selection of seven loci per clade, they provide preliminary evidence that the younger clades are more lowly methylated and display greater inter-individual methylation ranges (Faulk *et al.* 2013a). These clades are likely enriched for metastable epialleles.

Taking an unbiased approach, Oey and colleagues performed whole genome bisulphite sequencing on five A^{vy} mice and identified 356 regions showing inter-individual methylation variation, including 55 ERV-overlapping regions (Oey *et al.* 2015). Four of these appeared to experimentally validate. Perhaps the most valuable aspect of this study was the whole genome sequencing analysis of two A^{vy} mice with different coat colours. Only 32 single nucleotide variants were detected in coding sequences and no mutations were found near the A^{vy} allele, contesting arguments suggesting that the phenotypic variation observed among A^{vy} littermates is due to genetic variation.

The most recent genome-wide screen for metastable epialleles was conducted by our group. Anastasiya Kazachenka used whole genome bisulphite

sequencing (WGBS) datasets to identify variably methylated IAPs (VM-IAPs) in the B6 genome (Kazachenka *et al.* 2018). More than 50 of these have been experimentally validated. Much like the A^{vy} and $Axin^{Fu}$ alleles, VM-IAP methylation levels are variable between individuals but consistent across tissues within a single individual (**Figure 1.8A** and **B**). VM-IAPs are enriched for evolutionarily young IAP subclasses and are highly polymorphic between mouse strains. Approximately half of the identified candidates are solo LTRs of the IAPLTR2_Mm subclass and the other half is mostly made up of full-length elements with IAPLTR1_Mm flanking LTRs. Interestingly, the 5′ and 3′ LTRs of the same element behave independently with regards to their methylation state: for some, only one LTR displays variable methylation; for others, both LTRs are variable but their methylation levels are discordant. Inverse correlations are observed between VM-IAP methylation and adjacent gene expression for a subset of VM-IAPs, but these transcriptional effects are not widespread (Kazachenka *et al.* 2018). This indicates that *cis*-acting gene regulation is not a prerequisite for metastability. The term 'VM-IAP' will be used hereafter when discussing the metastable epialleles identified in this recent study screen to avoid confusion with the A^{vy} and $Axin^{Fu}$ loci.

Although the metastable epialleles discussed here are endogenous and naturally occurring, a chimeric long interspersed nuclear element (LINE) retrotransposon of the LINE-1 subclass was recently experimentally inserted into the *Axin* gene, inducing the kinked tail phenotype with variable penetrance much like the spontaneous $Axin^{Fu}$ insertion (Wang *et al.* 2019). Unlike $Axin^{Fu}$, the $Axin^{cL1}$ mutation is not associated with variable methylation levels but rather with variable H3K9ac enrichment. Additional experimental manipulations of this sort will be of great comparative value in determining the mechanisms underlying metastability.

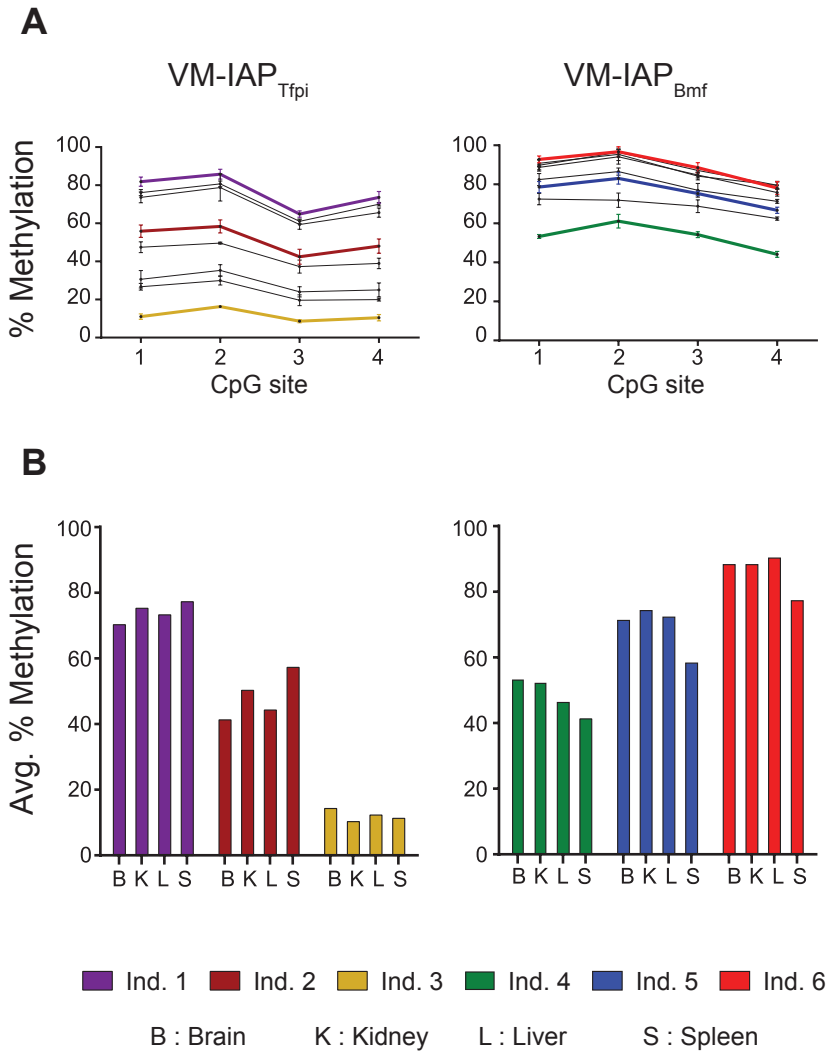


Figure 1.8: VM-IAP methylation levels are variable between individuals and constant across tissues. A. VM-IAP$_{Tfpi}$ and VM-IAP$_{Bmf}$ methylation levels assessed via bisulphite pyrosequencing in kidney tissues of eight individuals. Sequenced CpGs are the four most distal CpGs of the VM-IAP 5′ LTRs. Each line represents an individual. Colours are individual-specific and correspond to those used in panel B. Error bars represent the standard deviation between technical triplicates. **B.** VM-IAP$_{Tfpi}$ and VM-IAP$_{Bmf}$ methylation levels in brain (B), kidney (K), liver (L), and spleen (S) tissues. Methylation values are averaged across CpG sites. Individuals are colour-coded and correspond to those used in panel A. Figure adapted from Kazachenka *et al.* 2018.

1.4.2.2 Metastable epialleles in humans

The identification of human metastable epialleles is a challenging task due to the extensive genetic variation present in human populations. It is nevertheless an important endeavour in order to evaluate the extent to which they contribute to human phenotypic variation and to assess the relevance of using murine metastable epialleles as models to study epigenetic variation in humans. While analysing monozygotic (MZ) twin cohorts can aid in overcoming some of the challenges associated with genetic variation, a recent cautionary study described the phenomenon of "epigenetic supersimilarity" (ESS) between human MZ twins (Van Baak *et al.* 2018). Loci exhibiting ESS show greater epigenetic similarity between MZ twins than can be explained by their identical genomes, presumably due to the establishment of epigenetic marks prior to embryo splitting. The study highlights that MZ twins are not equivalent to isogenic mice that do not originate from the same zygote and suggests that ESS may contribute to non-genetic phenotypic similarities between MZ twins.

An alternative strategy to control for genetic differences in human populations has been to use a large number of genetically diverse methylomes. This approach has given rise to a growing list of putative human metastable epialleles, some of which are sensitive to environmental factors such as maternal nutrition and season of conception (Harris *et al.* 2013; Kessler *et al.* 2018; Silver *et al.* 2015; Waterland *et al.* 2010). The most recent of these studies argues that the definition of metastable epialleles should be relaxed to include variably methylated regions that are susceptible to genetic influence, at least in the human context where this issue is unavoidable (Kessler *et al.* 2018). Under this framework, additional sequence-dependent candidates identified in the past year can be added to the list of potential human metastable epialleles (Garg *et al.* 2018; Onuchic *et al.* 2018).

Interestingly, although IAP elements do not exist in humans, the bordering regions of putative human metastable epialleles appear to be enriched for TEs of the ERV and LINE families (Kessler *et al.* 2018; Silver *et al.* 2015). While some of the murine screens were specifically restricted to TEs based on the presence of IAPs at the A^{vy} and $Axin^{Fu}$ loci (Ekram *et al.* 2012; Faulk *et al.* 2013a; Kazachenka *et al.* 2018), others were unbiased and still found enrichment for repeats (Oey *et*

al. 2015; Weinhouse *et al.* 2011). Taken together, the mouse and human studies indicate that repeat elements play an important and conserved role in the establishment of inter-individual epigenetic variability. It is possible that metastable epialleles are a product of conflicting interactions between activating factors recruited to insertion sites and repeat repressive modifiers, an idea we will return to in section 1.4.4.3. This does not preclude, however, the possibility of metastable epialleles at unique non-repetitive regions maintained by distinct mechanisms. Although the human studies suggest these exist, a systematic screen for metastability at unique regions has not been conducted in the mouse.

1.4.3 Metastable epialleles as models of epigenetic inheritance

There is considerable interest in determining the extent to which epigenetic information can be passed on from one generation to the next, as this challenges the dogma dictating that heritable traits are strictly conferred by the sequence of DNA transmitted from parent to offspring. Epigenetic inheritance across generations has convincingly been shown to occur in a number of non-mammalian model organisms (Heard & Martienssen 2014; Miska & Ferguson-Smith 2016; Quadrana & Colot 2016). In contrast, this type of inheritance is rare in mammals due to the extensive genome-wide epigenetic reprogramming that takes place during mammalian development. Where it does occur, the driving mechanisms remain poorly understood. Metastable epialleles are frequently cited as the best example of this phenomenon in mammals.

1.4.3.1 Partial inheritance of parental epigenetic state

Epigenetic inheritance across generations is one of the most striking properties of A^{vy} and $Axin^{Fu}$ mice. In the case of A^{vy}, maternal (but not paternal) coat colour phenotype affects the range of phenotypes observed in the offspring; the coat colour distribution of offspring born to yellow mothers is shifted towards yellow compared to that of offspring born to pseudoagouti mothers (**Figure 1.9A**; Wolff 1978; Wolff *et al.* 1998). These experiments were conducted using inbred mouse strains, effectively eliminating the possibility of genetically mediated effects (Morgan *et al.* 1999; Wolff 1978; Wolff *et al.* 1998). In light of the

increased incidence of obesity observed in yellow mice, maternal inheritance of coat colour at A^{vy} was originally attributed to metabolic differences in the intrauterine environments of developing embryos (Wolff 1978). Elegant embryo transfer experiments showed that this is not the case. Transferring fertilised oocytes from yellow dams to black foster mothers not carrying the A^{vy} allele produces offspring with the same coat colour distribution as offspring born to yellow dams without embryonic intervention (Morgan *et al.* 1999). This confirms that the transmission of maternal coat colour to the next generation is an epigenetic process rather than an environmental one. The same study reported a grand-maternal effect at A^{vy}: transmission of the allele through two generations of pseudoagouti dams appeared to cause a greater shift towards pseudoagouti than transmission through a single generation (**Figure 1.9B**). It is unknown whether passage through a third or fourth generation of pseudoagouti females produces a further cumulative effect.

In contrast to A^{vy}, the $Axin^{Fu}$ allele exhibits epigenetic inheritance upon both maternal and paternal transmission. Parents with several tail kinks are more likely to produce offspring with kinked tails (Rakyan *et al.* 2003). Interestingly, the effect is more pronounced following paternal inheritance, at least in the 129P4/RrRk mouse strain (Rakyan *et al.* 2003). Of note, variable phenotypes and methylation states are reacquired in the next generation at both the A^{vy} and the $Axin^{Fu}$ loci despite the observed phenotypic bias in the F1 generation towards parental phenotype. More careful consideration must be given to the as-yet-unknown mechanism underlying this faithful reconstruction of variability from one generation to the next.

It is worth noting that all of the studies reporting A^{vy} or $Axin^{Fu}$ epigenetic inheritance use coat colour and tail morphology as phenotypic readouts of DNA methylation, respectively. No comprehensive statistical analysis has been conducted on parent and offspring DNA methylation data at the A^{vy} or $Axin^{Fu}$ loci, and very few have carried out offspring phenotyping in a manner blind to parental phenotype. While the correlation between IAP LTR methylation and phenotype severity is well documented for these classic loci, assessing the epigenetic inheritance at other metastable epialleles lacking visual phenotypes cannot rely on this form of classification.

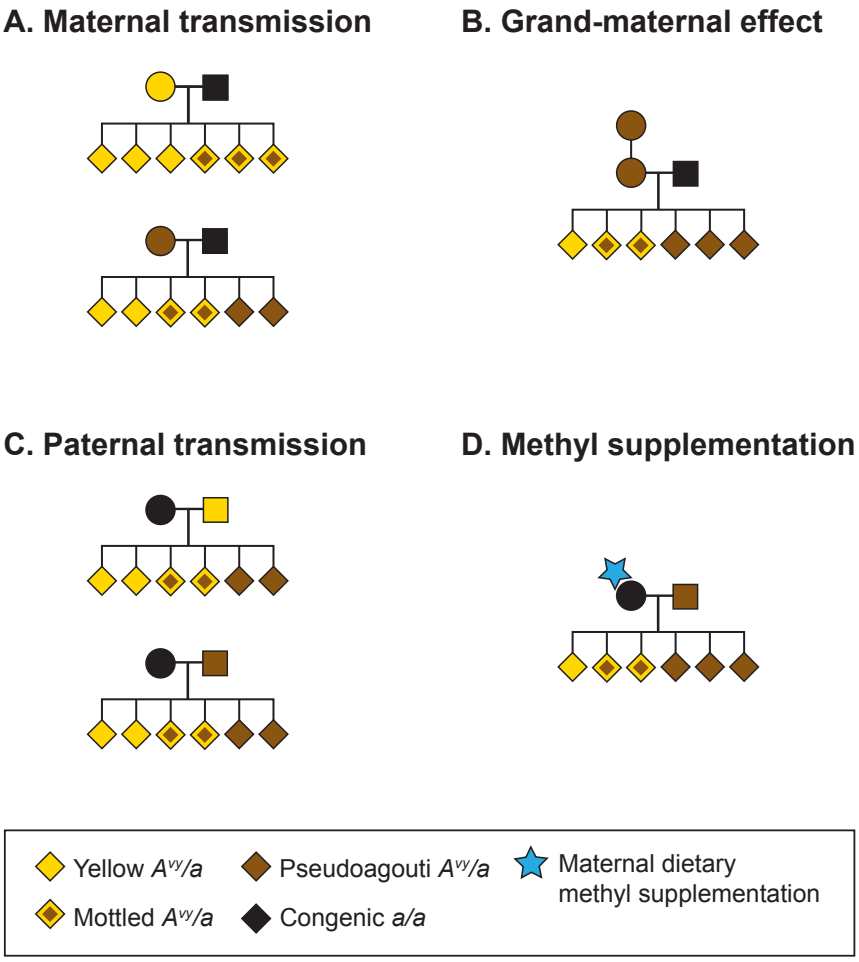


Figure 1.9: Inheritance patterns at the A^{vy} locus. A. Maternal coat colour phenotype influences the coat colour distribution in A^{vy}/a offspring in the B6 genetic background. **B.** A grand-maternal effect is observed when the A^{vy} allele is transmitted through two generations of pseudoagouti females, resulting in a more severe phenotypic shift. **C.** No inheritance is observed upon paternal transmission on this genetic background (adapted from Morgan *et al.* 1999). **D.** In utero dietary methyl supplementation shifts A^{vy} coat colour towards pseudoagouti (Waterland & Jirtle 2003). Pedigrees: circle—female; square—male; diamond—sex unspecified; a/a offspring are not included. In order to track the parental origin of the A^{vy} allele in these experiments, mice were bred to congenic *a/a* mice carrying the recessive *Agouti* null allele (*a*). Figure adapted from Bertozzi & Ferguson-Smith 2019.

1.4.3.2 Genetic background effects

The genetic background used to study the heritability of metastable epialleles is an important consideration, as both A^{vy} and $Axin^{Fu}$ inheritance patterns are influenced by the mouse strain the alleles are maintained on. The magnitude of maternal A^{vy} inheritance was shown to be dependent on the strain of the dam, with YS/ChWf females producing more pseudoagouti offspring than B6 or VY-Wf females (Wolff 1971, 1978). Similar genetic background effects have been reported for the penetrance of tail kink phenotypes associated with the $Axin^{Fu}$ allele (Ruvinsky & Agulnik 1990). Interestingly, when $Axin^{Fu}$/+ 129P4/RrRk male mice are crossed with A^{vy}/a B6 female mice, there is no paternal inheritance of tail phenotype, mimicking the inheritance pattern observed for A^{vy} coat colour on a B6 background (Morgan *et al.* 1999; Rakyan *et al.* 2003).

These dependencies on genetic background suggest that the A^{vy} and $Axin^{Fu}$ parent-of-origin effects are mediated by *trans* acting genetic factors. It is possible that genetic or cytoplasmic modifiers carried in B6 oocytes, but not 129P4/RrRk oocytes, promote complete epigenetic reprogramming of metastable epialleles, therefore preventing the transmission of paternal phenotype. Such strain-specific modification has been studied in other contexts but has not been tested at metastable epialleles (Latham & Sapienza 1998). It is nonetheless evident that genetic background has vital implications for the design and resulting generalizability of future experiments. We are once again reminded of the interdependence of genetic and epigenetic processes, and the problematic nature of investigating epigenetic variation in genetically heterogeneous contexts.

1.4.3.3 Developmental dynamics

It is difficult to reconcile genome-wide epigenetic reprogramming with the perpetuation of epigenetic states across generations. This has led to considerable debate in the ever-growing field of epigenetic inheritance. As mentioned above, some IAP elements are resistant to global demethylation in both the germline (Seisenberger *et al.* 2012) and during preimplantation development (Lane *et al.* 2003). This provides an attractive mechanism by which the IAP-driven A^{vy} and $Axin^{Fu}$ phenotypes could be inherited. Indeed, in sperm and oocytes, the

methylation levels of these alleles have been reported to reflect those observed in somatic tissues (**Figure 1.10**; Blewitt *et al.* 2006; Fernandez-Gonzalez *et al.* 2010; Rakyan *et al.* 2003). However, both the A^{vy} and $Axin^{Fu}$ loci are demethylated at the blastocyst stage, indicating that their associated IAPs are not resistant to the post-fertilisation wave of epigenetic reprogramming and demonstrating that DNA methylation is not perpetuated as the direct mediator of phenotypic heritability (**Figure 1.10**; Blewitt *et al.* 2006; Fernandez-Gonzalez *et al.* 2010). Little is known about the methylation dynamics at putative human metastable epialleles during early development although an analysis of human embryo methylomes suggests that the variable epigenetic states at these regions may be established during the gastrulation transition (Kessler *et al.* 2018).

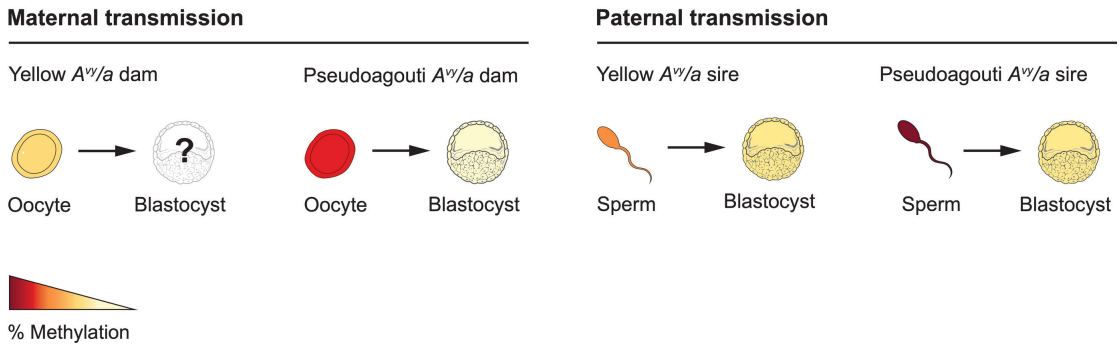


Figure 1.10: Post-fertilisation reprogramming of DNA methylation at the A^{vy} locus. DNA methylation levels of the A^{vy} allele in gametes and blastocysts upon maternal and paternal transmission are depicted based on clonal bisulphite sequencing data from Blewitt *et al.* 2006. Methylation levels in sperm and oocytes reflect the coat colour phenotype and somatic methylation levels of the individual, while blastocysts are largely unmethylated regardless of parental phenotype. This is consistent with erasure of DNA methylation during preimplantation stages. The methylation state of blastocysts produced by yellow A^{vy}/a dams has not been studied (depicted as a question mark). Figure from Bertozzi & Ferguson-Smith 2019.

1.4.4 Mechanistic insights into the establishment and maintenance of metastability

1.4.4.1 Histone modifications

The inter-individual methylation variation observed at metastable epialleles is likely associated with other variable epigenetic factors. Two studies have described the presence of variable histone marks at metastable epialleles, one conducted in A^{vy} liver tissue and the other in $Axin^{Fu}$ blastocysts. Mild enrichment of H3 and H4 di-acetylation was observed at the A^{vy} IAP LTR in yellow mice while H4K20me3 enrichment was detected in pseudoagouti mice. H4K20me3 is thought to be the most prominent histone modification at IAP LTRs, targeting them specifically over other types of repeats such as LINE-1 elements (Martens *et al.* 2005; Rangasamy 2013). No difference in H3K4me3 was found (Dolinoy *et al.* 2010). The study assessing the histone modification landscape at the $Axin^{Fu}$ locus at the blastocyst stage found significant differences in H3K4me2 and H3K9ac between blastocysts generated from penetrant and silent sires, suggesting histone marks may be involved in the transmission of tail phenotypes across generations (Fernandez-Gonzalez *et al.* 2010). Histone modifications have also been explored at VM-IAPs, but no consistent patterns have emerged other than an enrichment for the active marks H3K27ac, H3K9ac and H3K4me3 at the bordering regions of transcript-overlapping VM-IAPs (Kazachenka *et al.* 2018). A more in-depth characterization of metastable epiallele histone profiles, both in terms of the number of loci examined and the range of histone marks considered, will prove useful in assessing their role in establishing variable epigenetic states.

1.4.4.2 Drawing parallels between metastable epialleles and transgenes

The use of transgenic mice has been vital for the study of genome function and for the modelling of human disease. One of the challenges associated with producing mice carrying exogenous DNA constructs is the often-unpredictable cell-to-cell variability in transgene expression levels within a cell population or between individuals. The cause of such variegation remains poorly understood and has been attributed to a range of factors, including the repressive effects of

multi-copy transgene arrays, the proximity of the integration site to heterochromatin, and the presence of viral or plasmid-derived sequences within transgene constructs (Dobie *et al.* 1996; Festenstein *et al.* 1996; Garrick *et al.* 1998; Kjer-Nielsen *et al.* 1992; Martin & Whitelaw 1996).

Reminiscent of the properties of metastable epialleles, variegated transgenes are (1) linked to variable DNA methylation levels, (2) modulated by strain background, and (3) influenced by parental origin (Allen *et al.* 1990; Reik *et al.* 1987; Sapienza *et al.* 1987; Swain *et al.* 1987). For some transgenes, variable expressivity is recapitulated from one generation to the next in a predictable manner (Opsahl *et al.* 2002; Preis *et al.* 2003). Others exhibit memory of parental methylation level, their silenced state stably inherited to subsequent generations after passage through the germline (Allen *et al.* 1990; Hadchouel *et al.* 1987; Kearns *et al.* 2000; Sutherland *et al.* 2000). This heritable silencing is sometimes irreversible, and at other times reactivated upon transmission through the other parent or by crossing to a different strain.

Although fully heritable silencing is not a property of metastable epialleles, the overlapping characteristics with transgenes are worth considering while investigating the mechanisms underlying epigenetic stochasticity. Both are associated with foreign DNA sequences with regulatory potential, likely triggering similar host genome recognition and response pathways. Parallels have been drawn between transgenesis and retrotransposition before (Blewitt & Whitelaw 2013; Slotkin & Martienssen 2007); some have even classified variegated murine transgenes as metastable epialleles (Rakyan *et al.* 2002). In fact, a successful screen for modifiers of variegated transgenes, described below, confirms that transgenes and metastable epialleles share epigenetic modifiers.

Having made key contributions to our understanding of the unique molecular behaviour of the A^{vy}, $Axin^{Fu}$ and $Cabp^{IAP}$ loci, Emma Whitelaw and her team embarked on a large-scale *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis screen for modifiers of epigenetic variability (Blewitt *et al.* 2005; Daxinger *et al.* 2013). The study used a mouse line carrying a GFP reporter transgene expressed in a variegated fashion in red blood cells. Importantly, the variegated expression of this transgene is predictable: 55% of red blood cells express GFP in multi-generational isogenic mice. Offspring born to ENU-treated males were assessed

for enhancement or suppression of variegation by screening for shifts in the percentage of GFP-expressing red blood cells. The resulting mutations were designated *Modifiers of Murine Metastable Epialleles* (*Mommes*); dominant mutations were referred to as *MommeDs.*

Mapping the mutations associated with *MommeDs* is still ongoing, as is the characterisation of the affected genes (Ashe *et al.* 2008; Blewitt *et al.* 2005; Chong *et al.* 2007; Daxinger *et al.* 2012, 2013, 2016; Harten *et al.* 2014, 2015; Isbel *et al.* 2016; Sorolla *et al.* 2015; Youngson *et al.* 2011). More than 50 *MommeD* enhancers or suppressors of variegation have been identified in this screen. *MommeDs* that increase the proportion of GFP-expressing cells have mutations in genes acting as suppressors of variegation and involved in transgene silencing. Conversely, *MommeDs* resulting in a decrease in the proportion of GFP-expressing cells have mutations in genes that enhance variegation and promote transgene expression. Unsurprisingly, the majority of mutations haven fallen into genes with known epigenetic regulatory properties. These include genes involved in DNA methylation (e.g. *Dnmt1* and *Dnmt3b)*, histone modification (e.g. *Brd1*, *Hdac1, Setdb1, Trim28*), and chromatin remodelling (e.g. *Baz1b, Pbrm1, Smarca4, Smarca5*) (Blewitt & Whitelaw 2013). Previously unknown *D14Abb1e* was identified as a gene essential for early development (Harten *et al.* 2014). *D14Abb1e* is the mouse orthologue of *TASOR,* which was subsequently identified as a HUSH complex component following a similar screen for modifiers of PEV in human cells (Tchasovnikarova *et al.* 2015). In addition, many of the identified genes in the *MommeD* screen were previously detected in screens for modulators PEV in *Drosophila*, reflecting the highly conserved nature of epigenetic modifiers (Schotta *et al.* 2003).

In line with previously reported similarities between transgene and metastable epiallele epigenetic states, *MommeDs* were also found to modulate the A^{vy} locus. In particular, crossing *MommeD* heterozygotes with A^{vy}/a mice resulted in offspring with shifted coat colour distributions. In general, there was concordance between positive regulation of GFP-transgene expression and a shift in coat colour towards yellow, and vice versa. Yellow-shifting *MommeDs* include the mutants $Smchd1^{MD1}$, $Dnmt1^{MD2}$, $Setdb1^{MD13}$, and $Trim28^{MD9}$; pseudoagouti-shifting *MommeD*s include $Smarca5^{MD4}$, Rlf^{MD8}, and Wiz^{MD30}. These experiments notably identified paternal effect genes, whereby wild type pups

born to mutant sires exhibited changes in coat colour distribution. As a result, *Smarca5* and *Dnmt1* were the first ever reported paternal effect genes in the mouse (Chong *et al.* 2007). More recently, *Setdb1* was found to exhibit similar behaviour (Daxinger *et al.* 2016).

The genes underlying *MommeDs* have diverse functions at endogenous loci extending beyond the regulation of transgene variegation. For example, previously uncharacterized *Smchd1* has been shown to regulate long-range interactions on the inactive X chromosome and at Hox genes (Blewitt *et al.* 2008; Jansz *et al.* 2018). *Smchd1*^{MD1} mutants shift *A*^{vy} coat colour towards yellow upon maternal inheritance, but only in female offspring (Blewitt *et al.* 2005).

Complex interactions between the many genes uncovered by the *MommeD* mutagenesis screen are likely involved in the maintenance of epigenetic states at metastable epialleles. However, not all *MommeDs* have been studied with regards to their effect on *A*^{vy} coat colour, and, perhaps due to the nature of the screen, none affect the establishment of metastability itself. Considering that the HUSH complex spreads heterochromatin over inserted transgenes in mammalian cells and is implicated in the epigenetic regulation of murine retrotransposons (Robbez-Masson *et al.* 2018; Tchasovnikarova *et al.* 2015), further characterisation of HUSH-mediated repression may provide additional insight into the parallels between transgene and metastable epiallele regulation.

1.4.4.3 CTCF binding at VM-IAPs

CTCF has recently emerged as a potential regulator of metastable epialleles. CTCF is a multi-functional methyl-sensitive DNA binding protein, with 41% of cell-line-specific CTCF binding sites being associated with DNA methylation at unbound loci (Bell & Felsenfeld 2000; Phillips & Corces 2009; Wang *et al.* 2012). The methylation-dependent sites contain CpGs at specific positions in the DNA binding site (Wang *et al.* 2012). While CTCF binding at the *A*^{vy} and *Axin*^{Fu} loci has not been studied, CTCF is enriched at VM-IAPs compared to their methylation invariant counterparts in multiple tissues and across different developmental time points (Kazachenka *et al.* 2018). It is plausible that a molecular antagonism between repeat element silencing via DNA methylation and the maintenance of unmethylated CTCF binding sites is contributing to the stochastic establishment

of metastable epiallele methylation states (Kazachenka *et al.* 2018). In line with this model, Amir Hay in our group has demonstrated inverse relationships between VM-IAP methylation level and abundance of bound CTCF. Two recent studies in humans further support an association between CTCF and epigenetic variability: one finds an enrichment for CTCF binding sites at human metastable epialleles (Kessler *et al.* 2018) and the other implicates CTCF binding affinity in the regulation of stochastic switching between epigenetic states (Onuchic *et al.* 2018).

Despite these advances, the mechanisms underlying the establishment and maintenance of variable methylation levels at metastable epialleles remain poorly understood. The epigenetic pathways mediated by non-coding RNAs, KZFPs, and HUSH complex subunits are all candidate regulators of metastable epialleles.

1.4.5 Environmental modulation of metastable epialleles

1.4.5.1 Methyl supplementation in the A^{vy} mouse model

The A^{vy} mouse line has become a popular model for the study of environmentally induced epigenetic change, the most documented intervention being *in utero* methyl supplementation. Maternal dietary supplementation with methyl donors and co-factors, including folic acid, vitamin B12, choline, and anhydrous betaine, has been shown to shift A^{vy}/a offspring coat colour towards pseudoagouti (**Figure 1.9D**; Cooney *et al.* 2002; Cropley *et al.* 2006; Waterland & Jirtle 2003; Wolff *et al.* 1998). The shift in phenotype has been attributed to an increase in methylation at the A^{vy} IAP (Waterland & Jirtle 2003). However, having shown that the silent pseudoagouti version of the A^{vy} allele is not normally fully methylated but rather averages ~65% methylation, Cropley and colleagues compared the IAP methylation levels of methyl-exposed and unexposed pseudoagouti A^{vy}/a offspring and found no difference in methylation density at the silent IAP LTR. This suggests that the observed coat colour phenotypic change following *in utero* exposure to methyl donors is driven by an increase in methylation of the more lowly methylated allele (Cropley *et al.* 2010).

It is possible that the increased methyl donor availability is acting indirectly via substrates other than cytosine bases at the A^{vy} allele, but this has not been tested. Consistent with this finding, a study in wild-derived deer mice (which lack a repeat element at the *Agouti* locus) showed that Agouti-controlled pelt colour is susceptible to methyl donor supplementation in the absence of a variably methylated retroelement (Shorter *et al.* 2014).

Some of the previously discussed hallmark properties of metastable epialleles re-emerge in methyl supplementation studies. These include genetic background effects, whereby the magnitude of the A^{vy} coat colour shift is dependent on the mouse strain used for the experiment (Wolff *et al.* 1998). Additionally, the coat colour of offspring born to methyl-supplemented dams was only found to be altered when the A^{vy} allele was inherited paternally (Cropley *et al.* 2006). This is reminiscent of parent-of-origin effects observed in the absence of dietary supplementation (Morgan *et al.* 1999). Therefore, the fully reconstructed paternal allele may be more sensitive to modulation via methyl donor supplementation at this early embryonic stage. This does not rule out environmental sensitivity of the maternally inherited allele, since a subsequent study reported methyl supplement-induced alterations in offspring coat colour phenotypes upon maternal transmission (Waterland *et al.* 2007).

Studies investigating environmental modulation of the epigenome often consist of exposing dams for two weeks prior to breeding followed by maintenance of the experimental regimen throughout pregnancy and lactation. While this experimental design maximizes the chances of observing an effect, it limits mechanistic inferences that would otherwise be possible by narrowing the window of exposure to a specific developmental time point. For example, the confinement of methyl supplementation to a single week during mid-gestation (corresponding to primordial germ cell migration and epigenetic reprogramming) resulted in a shift in offspring coat colour (Cropley *et al.* 2006). Another study on A^{vy} mice showed that feeding A^{vy}/a offspring a methyl donor diet post-weaning for a period of 29 weeks neither shifts coat colour nor IAP LTR methylation levels (Warzak *et al.* 2015). Together, these studies reveal that early pre-implantation embryogenesis is not the only environmentally susceptible period in development yet confirms that coat colour phenotype in A^{vy} mice and its associated epigenetic control are fixed by the age of weaning. Further

experiments that fine-tune the exact period of environmental vulnerability will help identify the windows of opportunity and hence possible mechanisms contributing to changes in epigenetic state.

1.4.5.2 Innate versus induced epigenetic inheritance – a sense of semantics

Most studies on epigenetic inheritance across generations in mammals follow phenotypes or epigenetic changes triggered by ancestral exposures to environmental insults (e.g. Carone *et al.* 2010; Ng *et al.* 2010; Radford *et al.* 2014). Others track phenotypes in wild-type offspring caused by a mutation in a previous generation (e.g. Nelson *et al.* 2010; Padmanabhan *et al.* 2013). The volume of these studies is ever expanding. In response, it has become useful to distinguish *transgenerational* from *intergenerational* epigenetic inheritance. For true transgenerational epigenetic inheritance to take place, the induced phenotype must arise from germ cells never exposed to the original stimulus (Ferguson-Smith & Patti 2011; Heard & Martienssen 2014; Miska & Ferguson-Smith 2016). In the case of maternal exposure during pregnancy, the primordial germ cells of the developing embryo (the future F2 generation) are also exposed. The induced change must therefore persist at least to the F3 generation arising from unexposed germ cells. This is not an issue for paternal exposure, so the heritable effect can be considered transgenerational if it persists to the F2 generation, and intergenerational if it does not.

This nomenclature is confusing when applied to A^{vy} mice. Many have referred to the A^{vy} mouse model as one of the best lines of evidence for transgenerational epigenetic inheritance. This is based on the pivotal finding that maternal A^{vy} phenotype, which is epigenetically controlled, influences that of the offspring (Morgan *et al.* 1999). As mentioned previously, grand-maternal phenotype also affects A^{vy} coat colour (Morgan *et al.* 1999). The extent to which this compounding effect extends beyond the F2 generation is unclear. In addition, these effects occur naturally in the population and no environmentally or genetically triggered phenotype is being tracked across generations in these experiments. If it were to be reported that the coat colour of F0 females influences that of the F3 generation, regardless of F1 and F2 coat colours, then the term *transgenerational* could be used. To our knowledge, this has not been investigated. To control for

confounding F1 and F2 effects, such a study would require a large number of crosses extending down multiple generations. Therefore, it is currently unknown whether innate epigenetic inheritance at the A^{vy} locus is trans- or intergenerational.

That said, the unique non-genetic inheritance of the A^{vy} pelt patterns combined with their reported environmental susceptibilities have sparked interest in determining the heritability of environmentally induced epigenetic changes at the A^{vy} locus. *In utero* exposure to methyl donors was found to shift coat colour toward pseudoagouti in both the F1 and F2 generations without additional supplementation of F1 dams (Cropley *et al.* 2006). This implies that aspects of the mechanism of epigenetic change in response to exposure can persist (either directly or indirectly) throughout gamete maturation and embryo development. Whether or not the complete demethylation of A^{vy}/a embryos at the blastocyst stage occurs in this methyl-supplemented context is unknown. Given this finding, it follows that continuous methyl supplementation of F0, F1, and F2 dams might result in a cumulative phenotypic shift in offspring toward pseudoagouti. This hypothesis was tested but not substantiated (Waterland *et al.* 2007). However, a subsequent study showed that multi-generational methyl supplementation leads to a progressive increase in the proportion of pseudoagouti mice if the supplementation is coupled with selection for the silent pseudoagouti phenotype, whereby only pseudoagouti offspring are bred to produce the next generation (Cropley *et al.* 2012). This cumulative effect was reversed after discontinuing supplementation. Despite disagreements on the merits and shortfalls of studies on this topic (Cropley *et al.* 2007), it is clear that the effects of a maternal methyl supplementation do not extend beyond the F2 generation in the absence of continuous exposure. Hence, in the environmental (or induced) context, epigenetic inheritance at A^{vy} is intergenerational.

Thus, there is a need to discriminate between innate and induced epigenetic inheritance across generations. Equally important, however, is the distinction between generational and cellular (mitotic) epigenetic inheritance to differentiate parent-to-offspring and mitotic cell-to-cell transmission, respectively. The generational qualifier must therefore be kept. To avoid semantic headaches, the use of *transgenerational* and *intergenerational* should be reserved for cases where the phenotype is traced to a specific generation, and the

more generic term *generational* should be used otherwise. Therefore, instances of generational epigenetic inheritance are innate or induced and can be further categorised as inter- or transgenerational when appropriate (**Figure 1.11**). Considering the direction of studies in this field, the latter distinction will most often be reserved for induced contexts. Accordingly, A^{vy} mice display innate generational epigenetic inheritance and diet-induced intergenerational epigenetic inheritance.

1.4.5.3 Additional A^{vy}-influencing environmental exposures

Other environmental insults have been reported to influence coat colour in A^{vy} mice. Maternal ethanol consumption shifts offspring coat colour towards pseudoagouti, regardless of whether ethanol is administered preconceptionally or during gestation (Kaminen-Ahola *et al.* 2010). A shift in the same direction and an increase in methylation levels at the IAP LTR were observed following maternal supplementation of genistein, an isoflavone abundant in soy (Dolinoy *et al.* 2006). Intrauterine ionizing radiation has been reported to favour the silenced version of the A^{vy} allele in a dose- and sex-dependent manner, rescued by dietary anti-oxidants (Bernal *et al.* 2013). Maternal dietary bisphenol-A (BPA) consumption and lead exposure were independently shown to have the opposite effect on A^{vy}/a offspring coat colour distribution, shifting it towards yellow (Anderson *et al.* 2012; Dolinoy *et al.* 2007; Faulk *et al.* 2013b). More recently, maternal exposure to phthalates, commonly found in plastics and cosmetics, caused altered coat colour distributions and higher IAP LTR methylation levels in A^{vy}/a offspring (Neier *et al.* 2019). Finally, *in vitro* culture of zygotes to the blastocyst stage was found to significantly shift pup coat colour towards yellow and decrease IAP LTR methylation levels (Morgan *et al.* 2008).

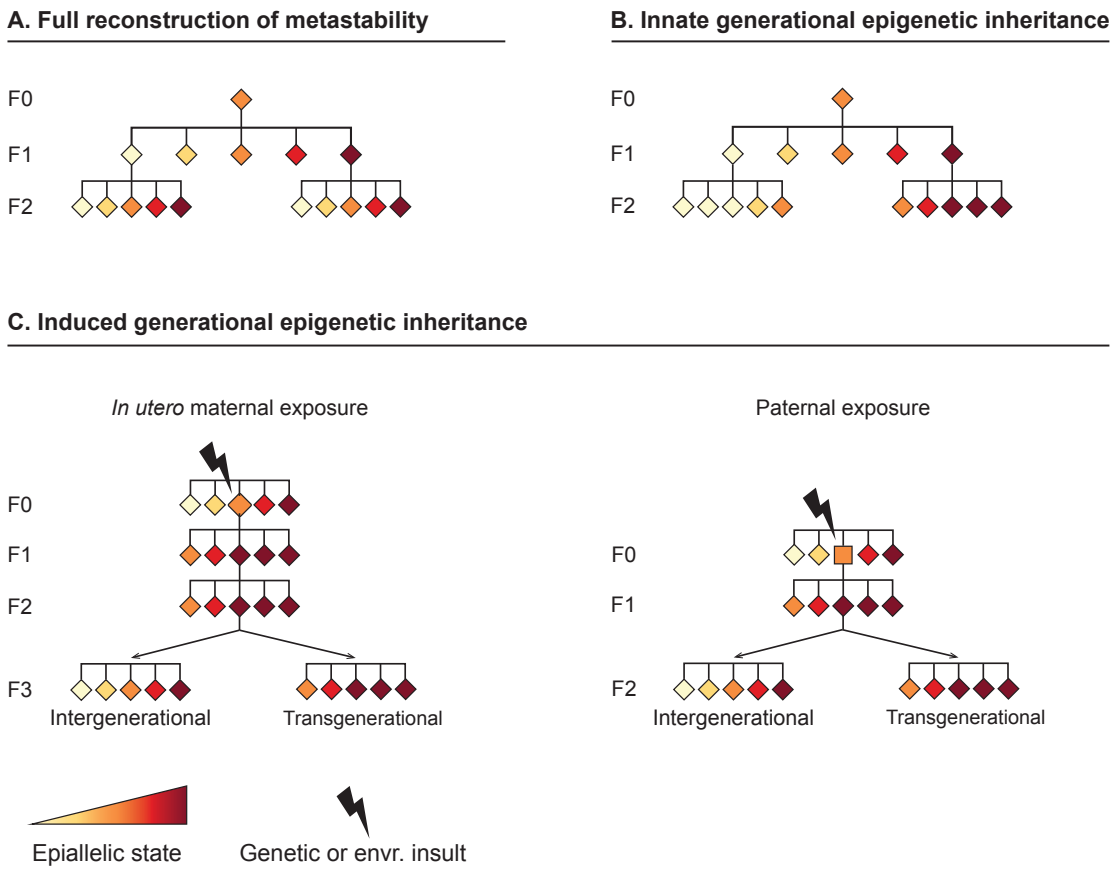


Figure 1.11: Reconstruction and heritability of epiallelic states. A. Reconstruction of epigenetic metastability. The full range of epiallelic states is reconstructed after passage through the germline, regardless of parental state. **B and C.** Generational epigenetic inheritance occurs when the parental epiallelic state influences that of the offspring. It can either be *innate*, occurring in the absence of an external trigger, or *induced*, defined by the cross-generational persistence of an epigenetic change brought about by a genetic or environmental insult inflicted in a previous generation. In the case of induced epigenetic inheritance, if the phenotypic or epigenetic perturbation persists at least to the F3 generation following *in utero* maternal exposure (or to the F2 generation following paternal exposure), the effect is *transgenerational*. It is *intergenerational* if it only persists to F2 (or F1 upon paternal transmission) (Ferguson-Smith & Patti 2011). Since the distinction between trans- and intergenerational inheritance is most often associated with induced instances, it is only depicted in the induced context in the diagrams above. Pedigrees: circle—female; square—male; diamond—sex unspecified. Figure adapted from Bertozzi & Ferguson-Smith 2019.

Research on A^{vy} environmental modulation has been controversial. A 2008 study on maternal consumption of casein and soy protein isolate, which contain genistein, showed no alteration of A^{vy}/a offspring coat colour (Badger *et al.* 2008). Similarly, an extensive analysis using generalised linear mixed models on a total of 426 mouse litters and six different dietary interventions was unable to reproduce previously reported effects of BPA and genistein on A^{vy}/a offspring coat colour (Rosenfeld *et al.* 2013). The same study revealed a strong parity effect, whereby changes in coat colour distribution were observed in offspring born from different parities within a single treatment group, highlighting the extreme care that must be taken in designing these experiments.

1.4.5.4 Environmental modulation of other loci

While most research programs on the environmental modulation of metastable epialleles have focused on A^{vy}, there is evidence that other loci are also susceptible. Maternal methyl supplementation causes a decrease in the incidence of kinked tails in $Axin^{Fu}$/+ offspring and an increase in DNA methylation at the $Axin^{Fu}$ locus (Waterland *et al.* 2006). Methylation levels at $Cabp^{IAP}$ are decreased in offspring born to BPA-exposed dams (Dolinoy *et al.* 2007), and mildly increased following lead exposure (Faulk *et al.* 2014). As observed for A^{vy}-associated phenotypes, *in vitro* culture of $Axin^{Fu}$/+ embryos from the zygote to the blastocyst stage leads to a more severe tail kink phenotype (Fernandez-Gonzalez *et al.* 2010).

Since the A^{vy} and $Axin^{Fu}$ loci arose from insertional mutations, commonly used laboratory mouse strains do not carry these loci. The recent identification of novel metastable epialleles in the B6 genome allows for the assessment of a repertoire of regions in the same set of environmentally perturbed mice to determine whether they respond synchronously, and to the same extent, to intrauterine environmental influences. One such study detected small tissue-specific DNA methylation differences at three variably methylated IAPs following perinatal lead exposure (Faulk *et al.* 2013b; Montrose *et al.* 2017). **Table 1.1** summarises the studies conducted to date concerning the environmental modulation of metastable epialleles.

Table 1.1: Environmental modulation of metastable epialleles.

In utero exposure	Locus	Effect on offspring		References
		Effect on phenotype	Effect on methylation	
Silencing				
Methyl donors	*A^vy^*	twd. pseudoagouti	Increase	Cooney *et al.* 2002; Wolff *et al.* 1998
	*Axin^Fu^*	twd. straight tail	Increase	Waterland *et al.* 2006
Genistein	*A^vy^*	twd. pseudoagouti	Increase	Dolinoy *et al.* 2006
	*A^vy^*	None	N/A	Rosenfeld *et al.* 2013
Casein and soy protein isolate[1]	*A^vy^*	None	N/A	Badger *et al.* 2008
Ethanol	*A^vy^*	twd. pseudoagouti	Increase	Kaminen-Ahola *et al.* 2010
Ionizing radiation	*A^vy^*	twd. pseudoagouti	Increase	Bernal *et al.* 2013
Dibutyl phthalate	*A^vy^*	twd. pseudoagouti	Increase	Neier *et al.* 2019
Activating				
BPA	*A^vy^*	twd. yellow[2]	Decrease	Dolinoy *et al.* 2007
	*A^vy^*	None	N/A	Rosenfeld *et al.* 2013
	*Cabp^IAP^*	N/A	Decrease	Dolinoy *et al.* 2007
Lead	*A^vy^*	twd. yellow	Cubic trend[3]	Faulk *et al.* 2013b
	*Cabp^IAP^*	N/A	Cubic trend[3]	Faulk *et al.* 2013b
	IAP 110	N/A	Decrease	Montrose *et al.* 2017
	IAP 236	N/A	Decrease	Montrose *et al.* 2017
	IAP 506	N/A	Decrease	Montrose *et al.* 2017
Embryo culture[4]	*A^vy^*	twd. yellow	Decrease	Morgan *et al.* 2008
	*Axin^Fu^*	twd. kinky tail	Decrease	Fernandez-Gonzalez *et al.* 2010

[1]Contains genistein
[2]Only following high levels of exposure
[3]Dose-dependent
[4]Not in utero

1.4.5.5 The A^{vy} mouse model: an epigenetic biosensor of environmental compromise?

The A^{vy} mouse model has been documented as a sensitive epigenetic biosensor of environmental compromise (Dolinoy 2008; Jirtle 2014; Waterland 2006). An ideal epigenetic biosensor is (1) particularly susceptible to a given environmental change and (2) exhibits an epigenetic response that is both predictable and easily detectable. Coat colour in A^{vy} mice appears to be acutely sensitive to slight changes in embryonic environment, likely via epigenetic influences at the A^{vy} locus. However, its innate epigenetic and phenotypic variability, established in large part by a stochastic process, is precisely what makes it a poor biological readout with little predictive value. The full range of coat colour phenotypes and associated methylation levels are observed in both control and exposed mice in these studies, requiring hundreds of mice to detect an effect and reach sufficient statistical power. Indeed, when over 2000 animals were analysed to assess the effect of *in utero* BPA and genistein exposures (both separately and together) on A^{vy} offspring coat colour, no significant shifts were observed (Rosenfeld *et al.* 2013). Together, this not only makes the use of the A^{vy} mouse model a costly and inefficient biosensor of environmental perturbation, but also questions its efficacy in this context. Further studies on the more recently identified metastable epialleles in B6 mice (Kazachenka *et al.* 2018) will clarify whether some regions are better biosensors than others, or whether perhaps metastable epialleles *en masse* can be used to build a multifactor epigenetic biosensor with enhanced predictive capabilities.

1.5 Research aims and thesis overview

1.5.1 Aims

Until recently, variable methylation of repeat elements in the absence of genetic variation was associated with a select few examples in mammals, mainly the A^{vy} and $Axin^{Fu}$ loci. These classic alleles have been the subject of intense investigation and public interest due to their striking phenotypes and unique biological characteristics, including epigenetic inheritance, environmental sensitivity, and unstable methylation states upon passage through the germline. The recently identified catalogue of VM-IAPs in the B6 genome allows these phenomena to be explored more comprehensively. Therefore, the aims of this project are as follows:

1. To characterise the inheritance patterns and developmental dynamics of VM-IAP methylation states;

2. To evaluate VM-IAP susceptibility to environmental modulation; and

3. To determine whether VM-IAPs are subject to parent-of-origin and genetic background effects that could potentially be harnessed to identify VM-IAP modifiers.

1.5.2 Structure and overview

The next five chapters contain the experimental results from this dissertation. Chapter 2 explores VM-IAP inheritance patterns, showing that stochastic methylation at VM-IAPs is re-established from one generation to the next regardless of maternal or paternal methylation. A breeding-intensive experiment reveals that epigenetic inheritance at metastable epialleles is rare, with only one out of six loci exhibiting memory of maternal methylation level. Chapter 3 delves into the developmental dynamics of VM-IAP methylation and presents evidence that all metastable epialleles, including A^{vy}, become

hypermethylated in the male germline by the rodent-specific DNA methyltransferase DNMT3C (Barau *et al.* 2016). Experiments on embryonic and placenta samples as well as on ESCs show that the mechanisms contributing to the establishment of variable methylation levels at VM-IAPs are active during preimplantation development.

Chapter 4 is devoted to investigating the susceptibility of VM-IAP methylation to various environmental contexts and to ageing. Results show that abnormal folate metabolism shifts VM-IAP methylation levels and influences VM-IAP-associated gene expression. In contrast, no significant effects are observed following exposure to the endocrine-disruptor bisphenol A (BPA) or to a maternal obesogenic diet. In addition, VM-IAP methylation is found to be stable across the murine lifespan. Together, the findings from this chapter suggest that VM-IAPs are selectively susceptible to altered environmental conditions.

Chapter 5 introduces genetic variation to the study of VM-IAPs. A reciprocal hybrid breeding scheme using B6 and CAST/EiJ (CAST) inbred mice uncovers parent-of-origin and genetic background effects at VM-IAPs. Subsequent backcrossing experiments demonstrate that these strain-specific effects can be exploited to identify genetic modifiers of VM-IAPs, opening up a new avenue of future investigation. Finally, Chapter 6 describes the unexpected identification of a recombination-induced polymorphic IAP in the B6 population associated with a unique epigenetic signature.

The main findings from this thesis are brought together and discussed in Chapter 7. In addition to proposing future experiments, the discussion conceptualises a model that places KZFP binding kinetics at the centre of the mechanism driving methylation stochasticity at VM-IAPs. The dissertation concludes with Chapter 8, which contains a detailed description of the methods used throughout.

Chapter 2

Inheritance of VM-IAP methylation states

2.1 Introduction and objectives

The A^{vy} and $Axin^{Fu}$ metastable epialleles are often cited as evidence that transgenerational epigenetic inheritance can occur in mammals in a genetically homogeneous context. Yellow A^{vy} dams maintained on a C57BL/6J (B6) genetic background produce more yellow offspring compared to pseudoagouti A^{vy} dams (Morgan *et al.* 1999). In $Axin^{Fu}$ mice maintained on a 129P4/RrRk genetic background, both maternal and paternal phenotypes influence the distribution of tail phenotypes in the offspring (Rakyan *et al.* 2003). These two well-characterised loci have established a paradigm for mammalian epigenetic inheritance, but it is unclear whether other genomic regions with similar genetic and epigenetic properties also display this behaviour.

The recent identification of variably methylated IAPs (VM-IAPs) in the B6 genome provides an opportunity to answer this open question. The experiments presented in this chapter explore the inheritance patterns observed at VM-IAPs and are driven by the following objectives:

1. To determine whether inter-individual methylation variation at VM-IAPs is maintained from one generation to the next;

2. To establish the degree to which metastable epialleles exhibit epigenetic inheritance.

2.2 Methodology

The methylation level of all VM-IAPs cannot be measured at the same time due to technological limitations. Thus, although more than 50 VM-IAPs have been experimentally validated in the B6 genome, a subset of 12 were selected for this project. All loci are not always tested in every experiment. **Table 2.1** contains information on the 12 VM-IAPs of interest, including their location, structure, subtype classification, and length. VM-IAPs are named based on their closest coding gene (placed in subscript), but this does not assume a functional relationship to said gene. In fact, there are cases where the closest annotated coding gene is more than 300 kb away from the VM-IAP.

The DNA methylation results presented in this thesis were generated by bisulphite pyrosequencing. The detailed procedure for this technique is described in the methodology chapter (Chapter 8) but the principles are described here to ensure comprehension of the data that follows. Genomic DNA is treated with bisulphite to convert all unmethylated cytosine (C) bases to thymine (T) bases, leaving methylated Cs intact. Bisulphite-converted DNA is subsequently used as template in a polymerase chain reaction (PCR) to amplify regions of interest, which are ideally ~300 bp in length. PCR products are sequenced via pyrosequencing, a live-tracking sequencing-by-synthesis technology that detects light emitted when pyrophosphate is released following the incorporation of a nucleotide (Tost & Gut 2007). The percent methylation at each CpG is quantified by calculating the ratio of incorporated Ts to incorporated Cs at each site (**Figure 2.1**). Bisulphite pyrosequencing is therefore more quantitative than traditional methods of site-specific DNA methylation quantification such as clonal bisulphite sequencing or Southern blotting.

Table 2.1: VM-IAPs of interest.

VM-IAP	Chr.[1]	Relative position	Structure	IAP subtype (LTR; internal region)	Length (bp)
VM-IAP$_{Mbnl1}$	3	Intragenic	Full-length	IAPLTR1_Mm; IAPEz-int	5,211
VM-IAP$_{Tfpi}$	2	Intragenic	Full-length	IAPLTR1_Mm; IAPEz-int	5,211
VM-IAP$_{Bmf}$	2	Intergenic	Truncated	IAPLTR2_Mm; IAPEz-int	3,609
VM-IAP$_{Rnf157}$	11	Intragenic	Full-length	IAPLTR1_Mm; IAPEz-int	5,304
VM-IAP$_{Eps8l1}$	7	Intragenic	Full-length	IAPLTR1_Mm; IAPEz-int	5,260
VM-IAP$_{Diap3}$	14	Intergenic	Solo LTR	IAPLTR2_Mm; N/A	478
VM-IAP$_{Fam78b}$	1	Intergenic	Solo LTR	IAPLTR2_Mm; N/A	471
VM-IAP$_{Marveld2}$	13	Intergenic	Full-length	IAPLTR1_Mm; IAPEz-int	5,273
VM-IAP$_{Rab6b}$	9	Intergenic	Solo LTR	IAPLTR2_Mm; N/A	510
VM-IAP$_{Sema6d}$	2	Intergenic	Solo LTR	IAPLTR2_Mm; N/A	471
VM-IAP$_{Gm13849}$	6	Intragenic	Full-length	IAPLTR1_Mm; IAPEz-int	5,267
VM-IAP$_{Slc15a2}$	16	Intragenic	Full-length	IAPLTR1_Mm; IAPEz-int	5,297

[1]Chromosome; complete genomic coordinates are listed in Appendix A, **Table A.1**.

The term 'VM-IAP methylation' is used in this thesis in reference to the methylation level of the four or five most distal CpGs of the VM-IAP 5′ LTRs. Although collecting methylation data from all of the CpGs lying in a given VM-IAP is of great interest, current technologies are not equipped to do this reliably. IAPs in the B6 genome display high sequence identity, so primers designed to target a specific IAP will also amplify non-specific products. To avoid this problem, either the forward or reverse PCR primer was designed to anneal to

the unique DNA that borders each VM-IAP, which limits methylation quantification to the distal CpGs of the LTR (**Figure 2.1**). Long read sequencing is expected to produce high quality and unambiguous methylation data for the internal portion of individual elements in the near future.

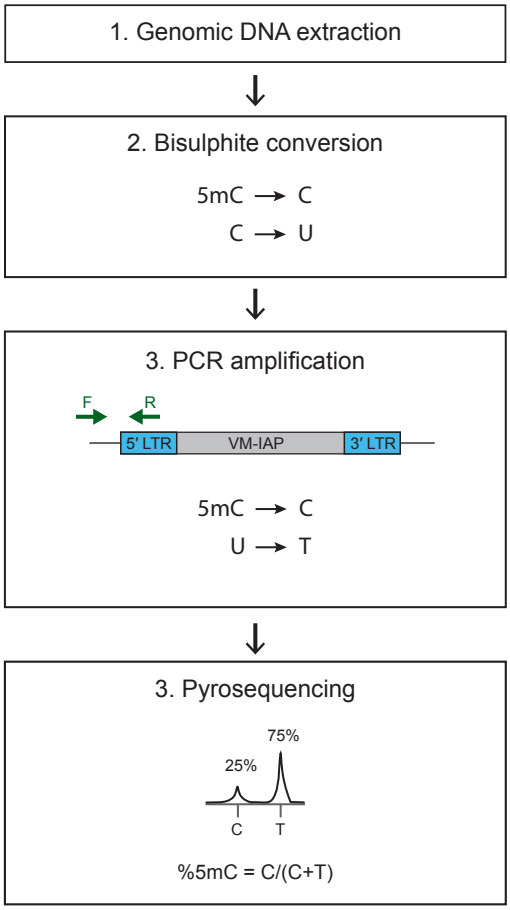


Figure 2.1: Bisulphite pyrosequencing workflow. Genomic DNA is bisulphite converted prior to PCR amplification of the distal region of the VM-IAP 5′ LTR. To avoid non-specific amplification, either the forward or reverse PCR primer (green arrows) is designed to target the region of unique DNA bordering the element. PCR products are sequenced using pyrosequencing technology and the percent methylation at each CpG is calculated as the ratio of incorporated Ts to incorporated Cs at each site.

2.3 Results

2.3.1 Characterisation of inter-individual methylation at VM-IAPs

2.3.1.1 VM-IAPs are metastable in ear tissue

VM-IAPs are a tissue-agnostic experimental system: they exhibit highly variable methylation levels across genetically identical individuals, but consistent methylation levels across tissues within an individual. Therefore, it is possible to ascertain somatic VM-IAP methylation levels using DNA extracted from ear notches rather than internal tissues. This approach to studying VM-IAPs is experimentally more efficient and allows mice to be kept alive, which will become an obvious benefit in the longitudinal ageing study discussed in Chapter 4. However, inter-individual methylation variation at VM-IAPs had to be verified in ear because the original validation experiments conducted on candidate VM-IAPs were carried out in brain, kidney, liver, and spleen samples (Kazachenka *et al.* 2018).

DNA was extracted from ear notch samples collected from B6 mice and VM-IAP methylation levels were quantified by bisulphite pyrosequencing. Each VM-IAP displayed inter-individual methylation variation that reflected the ranges previously observed in other tissues, indicating that it is appropriate to use ear DNA to study VM-IAP methylation dynamics (**Figure 2.2**; Kazachenka *et al.* 2018). While these 12 VM-IAPs will appear at various points throughout this thesis, only six out of the 12 (VM-IAP$_{Tfpi}$, VM-IAP$_{Mbnl1}$, VM-IAP$_{Bmf}$, VM-IAP$_{Marveld2}$, VM-IAP$_{Gm13849}$, and VM-IAP$_{Rnf157}$) are focused on for the rest of this chapter.

2.3.1.2 VM-IAP methylation is locus-specific within an individual

A qualitative comparison of methylation levels across VM-IAPs previously showed that within a single individual, one VM-IAP may be highly methylated, another lowly methylated, and yet another intermediately

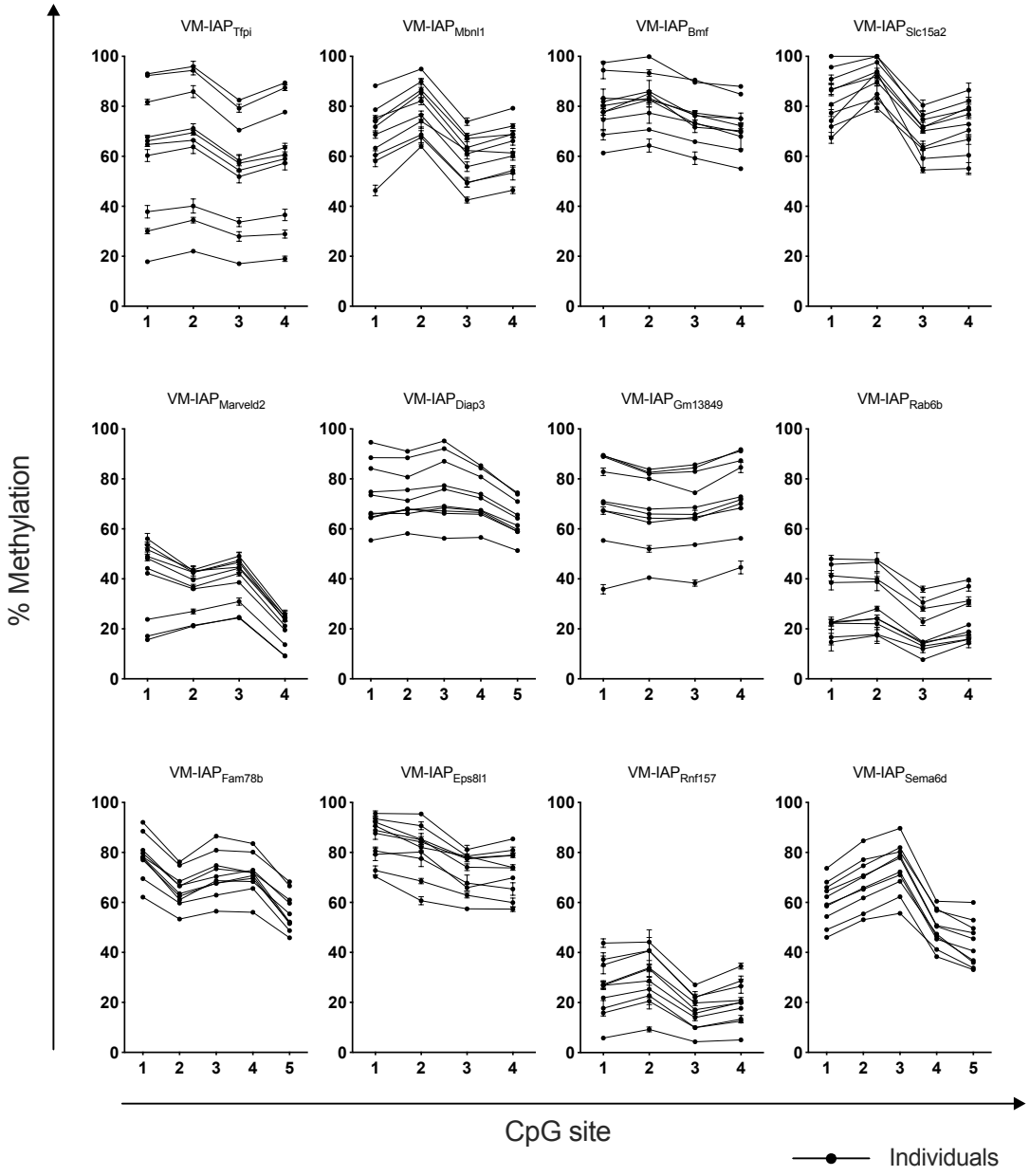


Figure 2.2: VM-IAPs exhibit inter-individual methylation variation in ear tissue. DNA was extracted from ear samples collected from B6 mice 10 days after birth (n = 10). Percent methylation was quantified for the 4 to 5 most distal CpGs of the VM-IAP 5′ LTRs via bisulphite pyrosequencing. Each line represents data from one individual and error bars represent the standard deviation between technical triplicates.

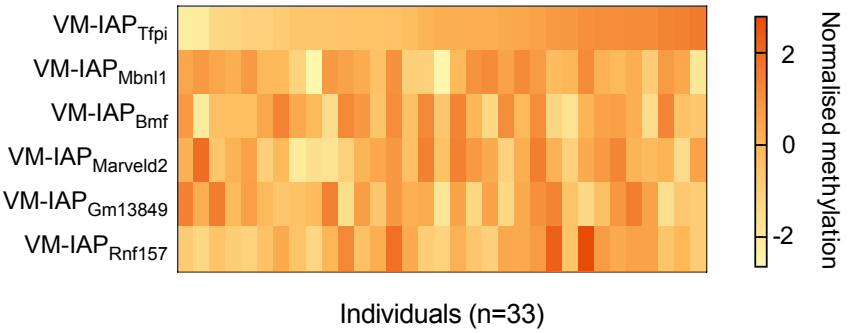
methylated (Kazachenka *et al.* 2018). To verify this finding and quantitatively explore co-variation of VM-IAP methylation levels within an individual, the methylation states of six VM-IAPs were assessed in 33 B6 mice. Methylation levels were averaged across the first four CpGs of the VM-IAP 5′ LTRs to obtain a single methylation percentage value for each individual at each VM-IAP. For ease of comparison across loci, these values were normalised to the methylation range of a given VM-IAP before generating a heatmap of relative methylation levels (**Figure 2.3A**). No clear co-variation patterns were observed. The normalised covariance matrix shown in **Figure 2.3B** indicates that the methylation level of each VM-IAP does not significantly correlate with that of other VM-IAPs within an individual, which confirms that methylation patterns are established independently at each VM-IAP in the same cellular context. Therefore, variable methylation patterns at these loci are not driven by a *trans*-acting individual-specific mechanism.

2.3.2 Cross-generational reconstruction of variable methylation and expression states at VM-IAPs

2.3.2.1 DNA methylation analysis in the F1 generation

To date, VM-IAPs have only been studied in mice from the same generational cohort. To determine cross-generational methylation dynamics, somatic VM-IAP methylation levels were assessed in pups born to highly or lowly methylated parents. For all tested VM-IAPs, the methylation variability between individuals observed at each locus in the F0 generation (**Figure 2.2**) was recapitulated in the F1 generation regardless of whether maternal (**Figure 2.4**) or paternal (**Figure 2.5**) methylation levels were high or low. This finding demonstrates that inter-individual methylation variation at VM-IAPs is faithfully reconstructed in the next generation after passage through the female and male germlines. Of note, the distinct range of methylation values observed at each region was also retained in the next generation. This disputes the possibility that the locus-specific ranges were the product of chance events

A



B

Covariance matrix

	VM-IAP$_{Tfpi}$	VM-IAP$_{Mbnl1}$	VM-IAP$_{Bmf}$	VM-IAP$_{Marveld2}$	VM-IAP$_{Gm13849}$	VM-IAP$_{Rnf157}$
VM-IAP$_{Tfpi}$	-	-0.087[1] 0.630[2]	0.005 0.980	0.061 0.737	-0.263 0.140	0.360 0.040
VM-IAP$_{Mbnl1}$	-0.087[1] 0.630[2]	-	0.092 0.609	0.233 0.192	0.214 0.231	0.316 0.073
VM-IAP$_{Bmf}$	0.005 0.980	0.092 0.609	-	0.148 0.413	-0.139 0.439	0.211 0.238
VM-IAP$_{Marveld2}$	0.061 0.737	0.233 0.192	0.148 0.413	-	0.215 0.229	0.071 0.693
VM-IAP$_{Gm13849}$	-0.263 0.140	0.214 0.231	-0.139 0.439	0.215 0.229	-	0.048 0.792
VM-IAP$_{Rnf157}$	0.360 0.040	0.316 0.073	0.211 0.238	0.071 0.693	0.048 0.792	-

[1]Pearson correlation coefficient
[2]P-value (Bonferroni-adjusted α value = 0.003)

Figure 2.3: Covariance analysis reveals that VM-IAP methylation states are locus-specific. **A.** Heatmap comparing the normalised methylation levels of six VM-IAPs in 33 B6 mice. Values were normalised to the inter-individual methylation range of each locus. Individuals are sorted based on normalised VM-IAP$_{Tfpi}$ methylation levels. **B.** Covariance matrix reveals no statistically significant associations between loci within an individual. Pearson correlation coefficients range from -1 to 1; values close to 0 specify independence of two parameters. P-values indicate the likelihood of obtaining each correlation coefficient by chance.

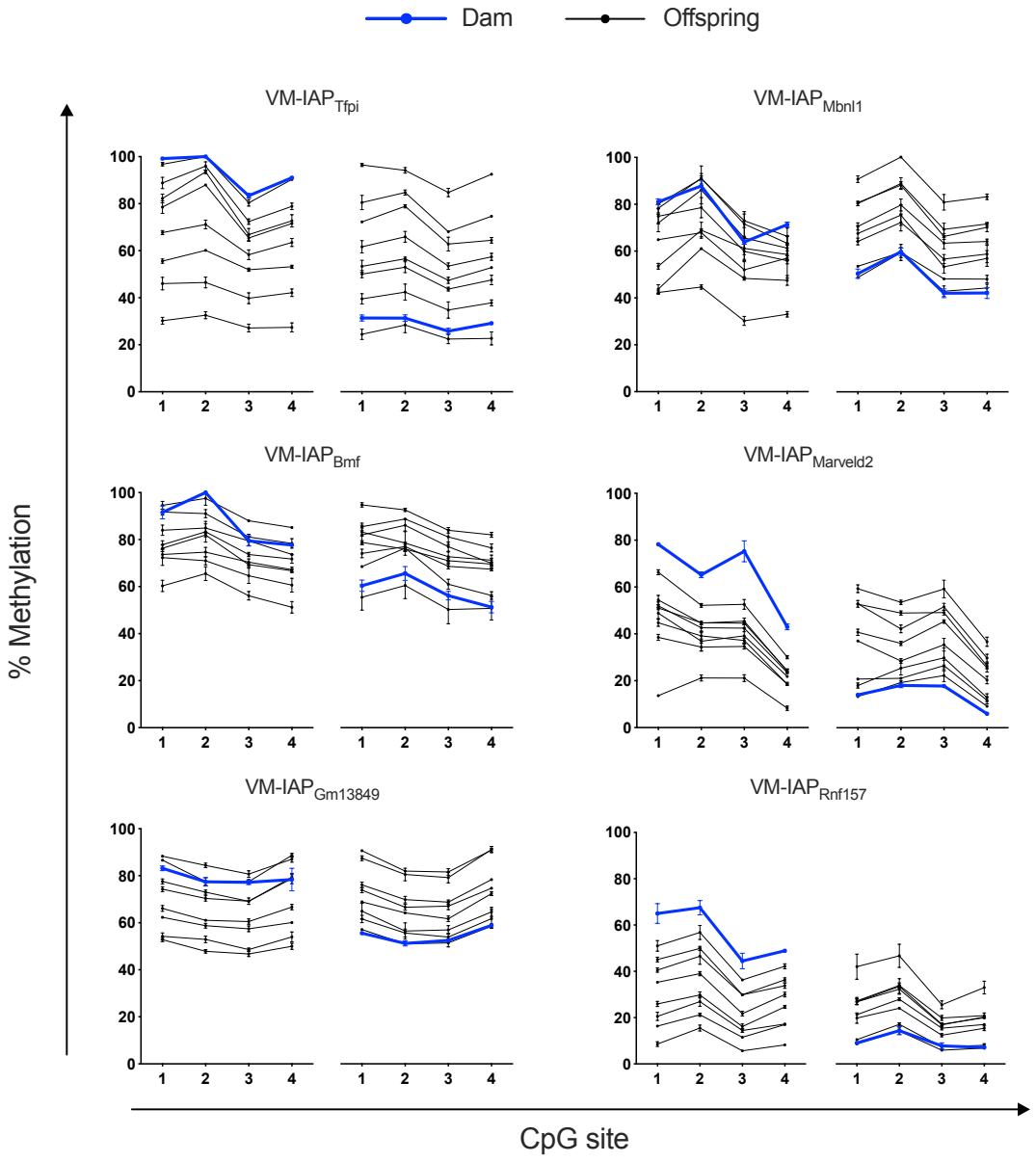


Figure 2.4: VM-IAPs are reconstructed as variable loci in the next generation regardless of maternal methylation level. Female B6 mice were assessed for methylation via bisulphite pyrosequencing at six VM-IAPs and those showing high or low methylation levels for at least one VM-IAP were set up for breeding. Each line represents data from one individual and error bars show the standard deviation between technical triplicates. Eight pups were analysed per dam. Both maternal and offspring methylation levels were quantified using ear notch DNA.

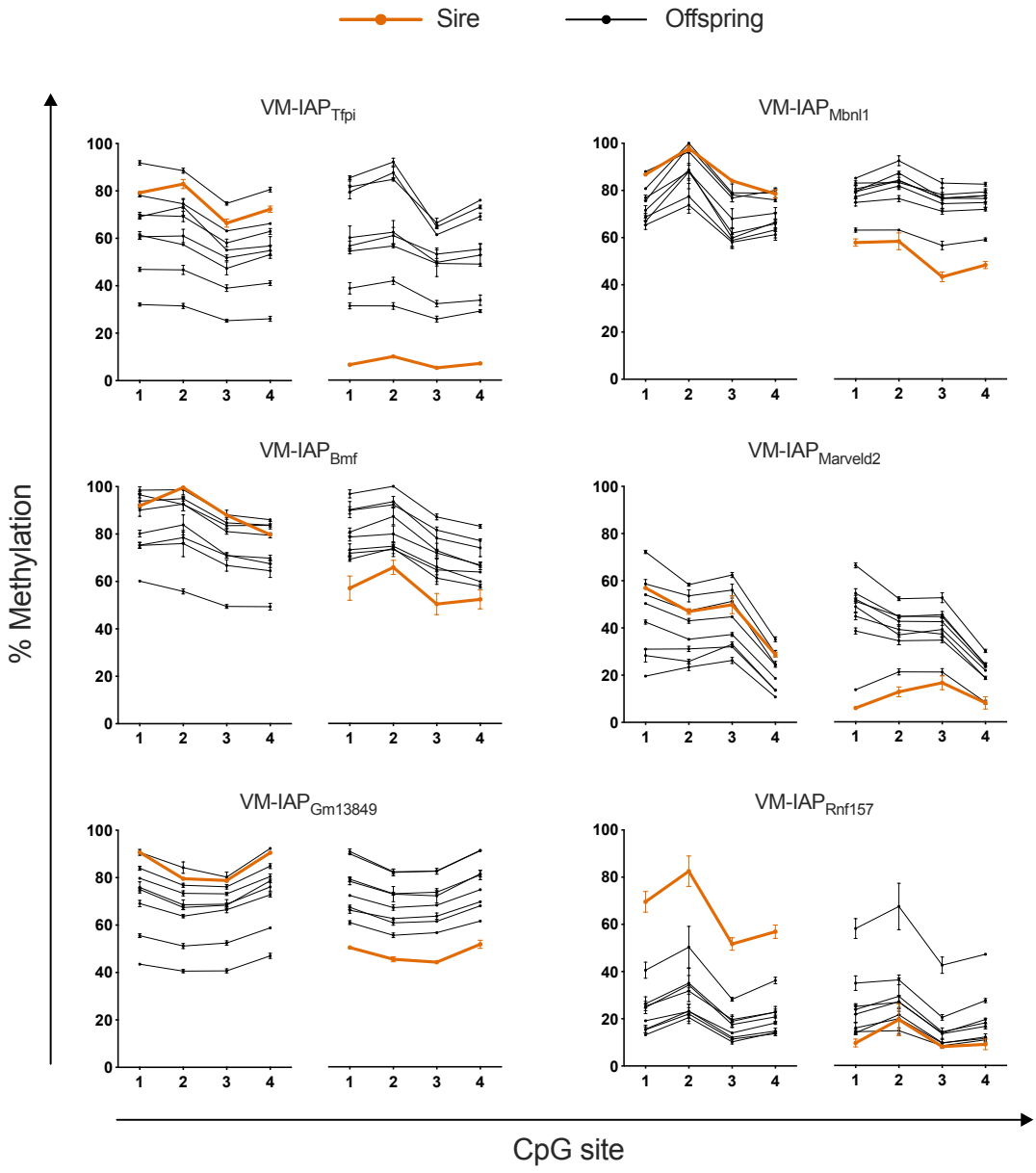


Figure 2.5: VM-IAPs are reconstructed as variable loci in the next generation regardless of paternal methylation level. Male B6 mice were assessed for methylation via bisulphite pyrosequencing at six VM-IAPs and those showing high or low methylation levels for at least one VM-IAP were set up for breeding. Each line represents data from one individual and error bars show the standard deviation between technical triplicates. Eight pups were analysed per sire. Both paternal and offspring methylation levels were quantified using ear notch DNA.

during the random selection of experimental mice. Instead, there seems to be a mechanism that drives stochastic methylation bounded between fixed locus-specific maxima and minima. This mechanism remains poorly understood and likely involves a combination of element-specific factors (e.g. the DNA sequence and the proteins that bind to it) and insertion-site-specific factors (e.g. the genomic context and chromatin architecture). A large portion of the discussion in Chapter 7 is devoted to proposing a model for this process and incorporates the knowledge acquired in this chapter and others.

Underlying genetic variation is always a potential explanation for putative epigenetic phenomena. Prior to the finding explained above, it remained a possibility that the variable methylation levels observed across inbred B6 mice at VM-IAPs were brought about by spontaneous DNA mutations. It is feasible that a single nucleotide polymorphism (SNP) in the B6 population, present in some mice but not in others, could lead to variable methylation at retrotransposons. If this were the case, all offspring born to a single B6 mouse would inherit the hypothetical mutation and would only display the associated parental methylation state. Because a single B6 mouse was found to produce offspring that exhibit the full methylation range observed at a given VM-IAP, we can assert that VM-IAP methylation levels are not genetically conferred. This does not preclude a genetic link to the establishment of metastability itself, as will become important in Chapter 5 when VM-IAPs are examined in the context of reciprocal hybrids and strain-specific modifiers.

2.3.2.2 Adjacent gene expression analysis in the F1 generation

In a manner analogous to the A^{vy} and $Axin^{Fu}$ loci, a subset of VM-IAPs show inverse correlations between their methylation state and the expression of adjacent genes. This occurs in splenic tissues at VM-IAP$_{Slc15a2}$ and VM-IAP$_{Eps8l1}$ (Kazachenka *et al.* 2018). To determine whether this relationship is sustained in the next generation, genomic DNA and total RNA were extracted from maternal and offspring spleen samples. VM-IAP$_{Slc15a2}$ and VM-IAP$_{Eps8l1}$ methylation levels were quantified via bisulphite pyrosequencing and *Slc15a2* and *Eps8l1* expression levels were assessed via quantitative real-time PCR

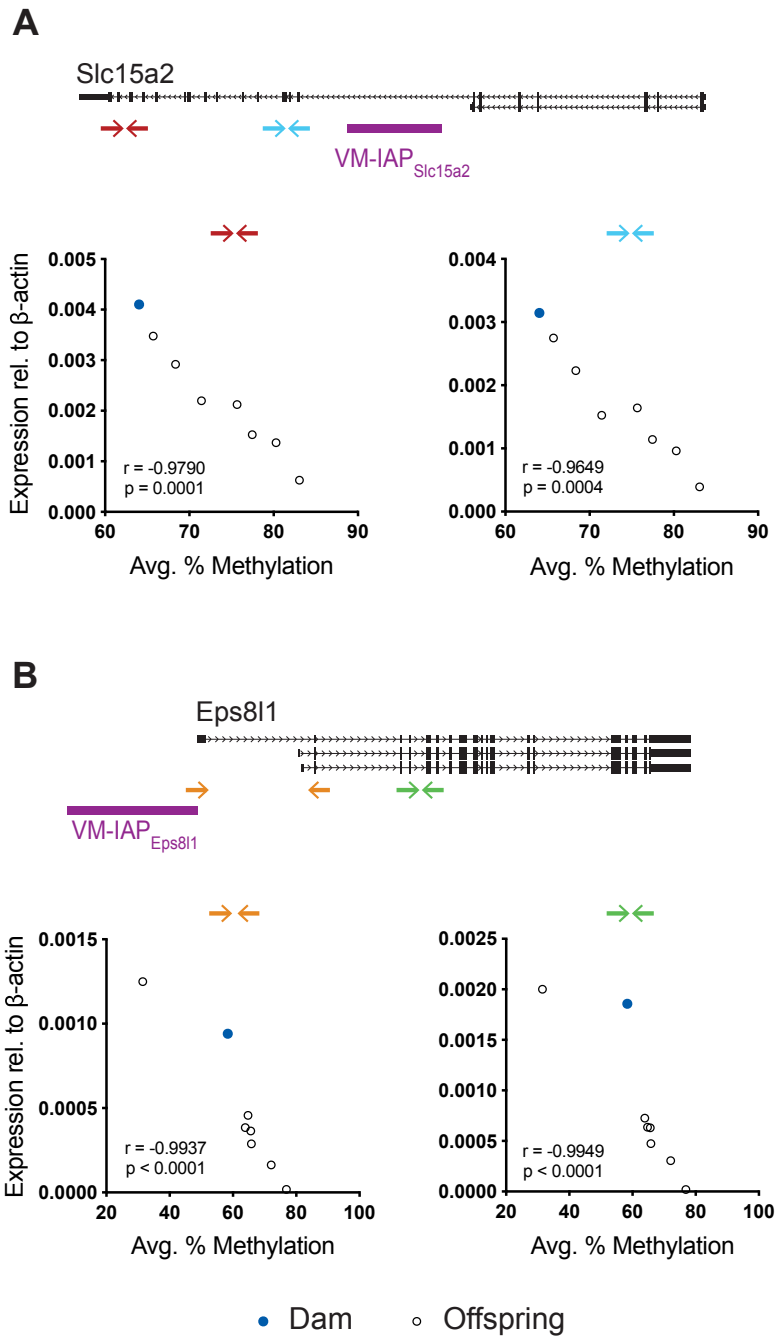


Figure 2.6: Inverse correlations between neighbouring gene expression and VM-IAP methylation are recapitulated in the F1 generation. *Slc15a2* (**A**) and *Eps8l1* (**B**) expression levels are inversely correlated with VM-IAP$_{Slc15a2}$ and VM-IAP$_{Eps8l1}$ methylation levels in the F1 generation, respectively (two-tailed Pearson excluding maternal data point; r, Pearson correlation coefficient). Diagrams of VM-IAPs in relation to their neighbouring gene are depicted above the graphs (drawn to scale). Gene transcripts from the University of California, Santa Cruz (UCSC) Genome Browser (Haeussler *et al.* 2019) are shown in black and VM-IAPs in purple. Colour-coded arrows represent the location of qRT-PCR primers. Two different primer pairs that span exon-exon junctions were used for each gene. Expression was quantified in spleen tissue and shown relative to the housekeeping gene *β-Actin*. Hollow black circles represent F1 individuals. Blue circles represent maternal F0 expression.

(qRT-PCR) using two different primer pairs for each gene. As previously observed in the F0 generation (Kazachenka 2018), expression levels of both genes were variable among F1 littermates and inversely correlated with VM-IAP methylation levels (**Figure 2.6**). Together with the results from the previous section, these findings indicate that the unique methylation signature of each VM-IAP and its effect on transcription are re-established from generation to generation.

2.3.3 Epigenetic inheritance is rare at VM-IAPs

2.3.3.1 VM-IAP$_{Gm13849}$ exhibits weak maternal inheritance

Section 2.3.2.1 showed that a single B6 mouse has the capacity to produce offspring that exhibit the entire range of possible methylation values for a given VM-IAP. However, this finding does not prohibit a bias in offspring methylation levels towards parental methylation level. This is illustrated by the A^{vy} locus: a pseudoagouti A^{vy}/a dam is capable of giving birth to yellow, mottled, and pseudoagouti offspring, but is more likely to produce pseudoagouti offspring than a yellow A^{vy}/a dam is. A litter of eight pups, as used in the experiments described above, is not a large enough sample size to detect such skews. This is particularly true if the effect sizes are small. Therefore, a breeding-intensive experiment was conducted on VM-IAPs to assess the prevalence of epigenetic inheritance at metastable epialleles.

The breeding schemes used for the inheritance studies on the A^{vy} locus consisted of reciprocal crosses between A^{vy}/a and a/a mice. Similarly, those conducted on the $Axin^{Fu}$ locus involved crossing $Axin^{Fu}$/+ mice to +/+ mice. This enabled the researchers to track the parental origin of the allele, with half the offspring inheriting a single copy and half inheriting none. These insertional mutations did not originally arise in the mouse strain on which they are currently maintained, allowing the inheritance analyses to be carried out in a genetically homogenous context apart from the allele in question. This experimental design is not applicable in our case because VM-IAPs were identified in the B6 genome. One alternative was to cross B6 mice to another

inbred mouse strain that lacked the VM-IAPs in question, but this option was rejected because VM-IAPs are highly susceptible to genetic background effects (described at length in Chapter 5). Therefore, intercrosses were set up between B6 mice, and maternal and paternal contributions to offspring methylation level were analysed using linear mixed-effects models (LMMs).

VM-IAP 5′ LTR methylation levels were quantified in ear samples collected from adult breeding pairs and their offspring (**Figure 2.7**). LLMs allowed for the incorporation of breeding pairs and litters as random effects, thereby controlling for the non-independence of siblings and littermates, respectively. Maternal methylation level, paternal methylation level, and offspring sex were included as potentially predictive fixed effects. Methylation levels were averaged across the four distal CpGs of the 5′ LTR for each individual and subjected to a logit transformation (log of the odds ratio) before feeding into the model. Logit transformations are commonly applied when working with discrete data such as percentages. Given the locus-specificity of VM-IAP methylation levels (**Figure 2.3**), a separate model was generated for each VM-IAP. Models were built using the lmer() function from the lme4 package in R (Bates *et al.* 2015). Effect sizes were quantified by computing marginal R^2 values for each fixed effect using the r2beta() function from the r2glmm package in R (Jaeger *et al.* 2017).

For five out of the six VM-IAPs tested, neither maternal nor paternal methylation level had a significant effect on offspring methylation levels (**Table 2.2**). For VM-IAP$_{Gm13849}$, maternal methylation level significantly affected offspring methylation levels ($p = 0.004$; $q = 0.036$; **Table 2.2**). In contrast, paternal methylation level did not have a significant effect on offspring methylation levels. This difference in heritability between parental lineages is consistent with the pattern observed for A^{vy} on a B6 background. Although the maternal effect at this locus was significant, the marginal R^2 value was small, which indicates that maternal methylation only accounts for a small fraction of the methylation variability observed in the next generation ($R^2 = 0.166$).

Given that a single VM-IAP displayed evidence of epigenetic inheritance, a smaller-scale experiment on VM-IAP$_{Gm13849}$ was designed to validate this

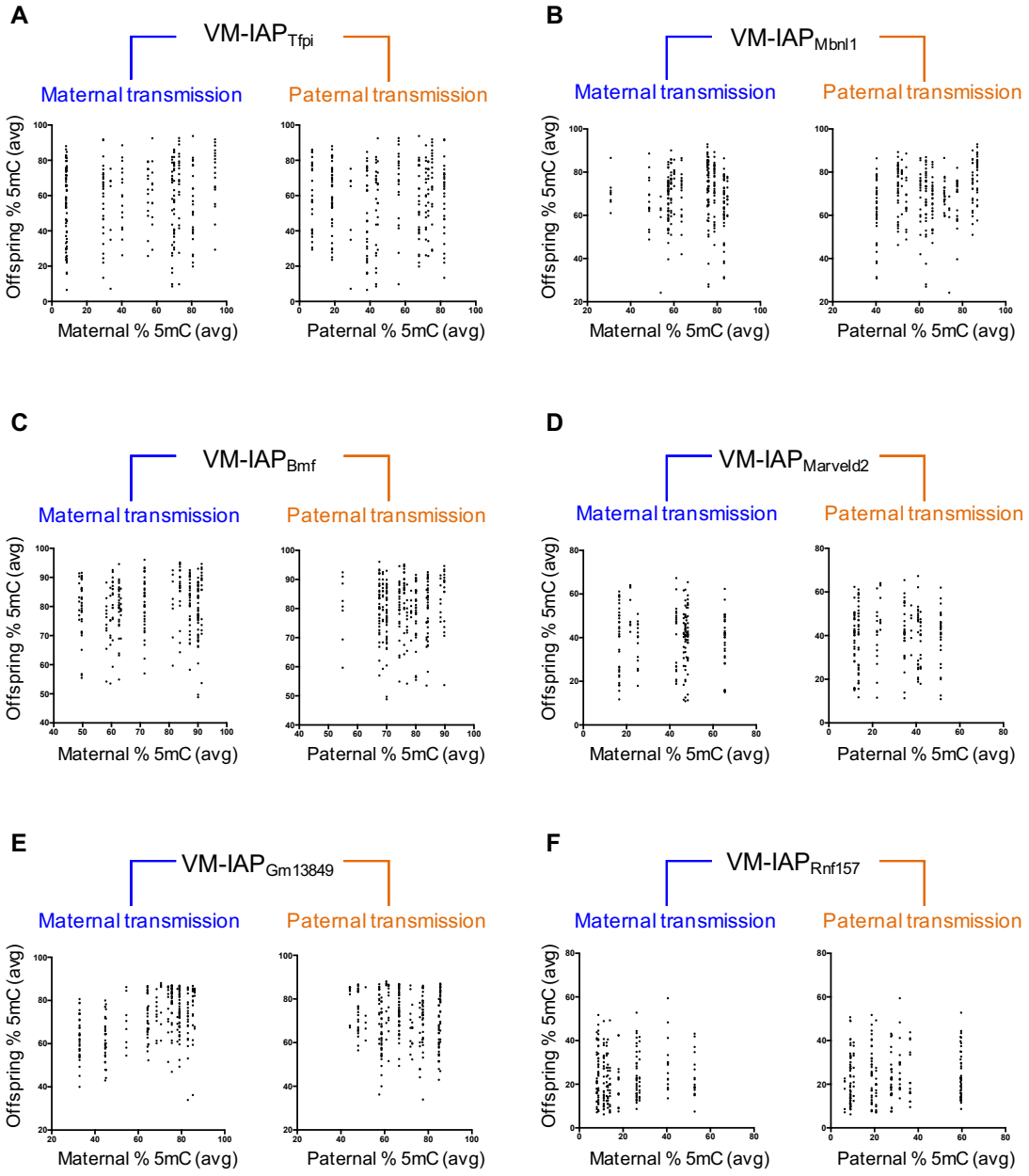


Figure 2.7: Inheritance analysis at VM-IAPs. A–F. Methylation levels at VM-IAPs were quantified from ear samples taken from B6 breeding pairs and their offspring. Offspring methylation level is plotted against maternal and paternal methylation level separately for VM-IAP$_{Tfpi}$ (A), VM-IAP$_{Mbnl1}$ (B), VM-IAP$_{Bmf}$ (C), VM-IAP$_{Marveld2}$ (D), VM-IAP$_{Gm13849}$ (E), and VM-IAP$_{Rnf157}$ (F). Each dot represents the methylation level of a single F1 individual. Methylation levels represent averages across the first four distal CpG sites of the VM-IAP 5′ LTRs. LMM summary statistics highlighting VM-IAP$_{Gm13849}$ as the sole locus that exhibits maternal heritability can be found in Table 2.2.

Table 2.2: Linear mixed-effects models: summary statistics for fixed effects.

	Estimate	Standard error	t value	R^{2*}	p-value	q-value
VM-IAP$_{Tfpi}$ (n = 264)						
Maternal % 5mC†	0.094	0.062	1.52	0.02	0.171	0.513
Paternal % 5mC†	0.025	0.077	0.32	0.001	0.757	0.847
Sex	0.152	0.121	1.26	0.006	0.21	0.54
VM-IAP$_{Mbnl1}$ (n = 265)						
Maternal % 5mC†	0.117	0.114	1.03	0.012	0.323	0.569
Paternal % 5mC†	0.241	0.099	2.43	0.068	0.048	0.173
Sex	-0.039	0.065	-0.6	0.001	0.547	0.757
VM-IAP$_{Bmf}$ (n = 278)						
Maternal % 5mC†	0.119	0.096	1.24	0.021	0.24	0.54
Paternal % 5mC†	-0.038	0.152	-0.25	0.001	0.8	0.847
Sex	0.071	0.07	1.01	0.003	0.31	0.569
VM-IAP$_{Marveld2}$ (n = 183)						
Maternal % 5mC†	-0.016	0.085	-0.19	0	0.849	0.849
Paternal % 5mC†	0.023	0.081	0.29	0.001	0.776	0.847
Sex	0.204	0.088	2.32	0.026	0.022	0.113
VM-IAP$_{Gm13849}$ (n = 262)						
Maternal % 5mC†	0.301	0.08	3.78	0.166	0.004	0.036
Paternal % 5mC†	-0.091	0.098	-0.92	0.011	0.379	0.569
Sex	0.142	0.063	2.26	0.017	0.025	0.113
VM-IAP$_{Rnf157}$ (n = 213)						
Maternal % 5mC†	0.032	0.093	0.35	0.001	0.73	0.847
Paternal % 5mC†	0.068	0.073	0.94	0.008	0.36	0.569
Sex	0.367	0.082	4.49	0.078	1.20E-05	2.16E-04

*R^2: semi-partial R^2 for each fixed effect.
†Averaged across CpGs and logit-transformed.

finding in a separate set of individuals. Five highly methylated and five lowly methylated B6 females were selected for breeding and VM-IAP$_{Gm13849}$ methylation levels were measured in first-litter offspring. The methylation level of offspring born to highly methylated mothers was significantly higher than that of offspring born to lowly methylated mothers, which corroborates the previous LMM result (p = 0.0069; n = 5 litters per group; **Figure 2.8**). Together, these studies indicate that inheritance of methylation levels is not a universal feature of metastable epialleles. Instead, VM-IAPs show evidence of epigenetic reprogramming and faithful re-establishment of metastability from one generation to the next regardless of parental epigenetic state.

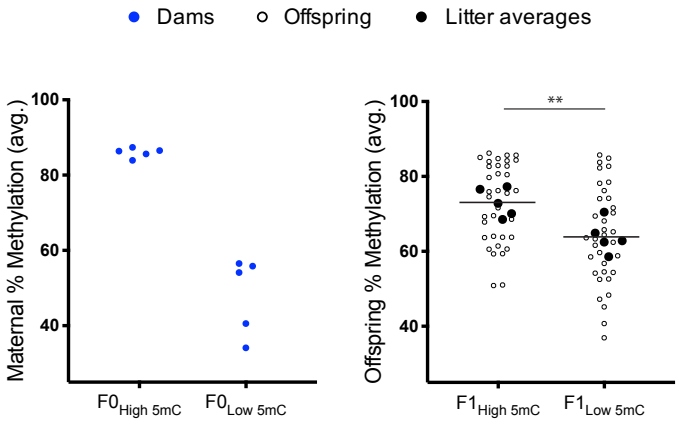


Figure 2.8: Independent validation of the maternal heritability effect observed at VM-IAP$_{Gm13849}$. B6 females were epigenotyped for methylation at VM-IAP$_{Gm13849}$ and only highly or lowly methylated females were selected for breeding (blue circles). The methylation level of all first-litter offspring was quantified (hollow black circles) and litter averages (black circles) were used for statistical analysis to ensure datapoint independence (one-sided t test; p-value: 0.0069; n = 5 litters per group). Black lines represent litter means. All data were generated from ear notch samples.

2.3.3.2 VM-IAP$_{Rnf157}$ methylation is modulated by sex

Using LMMs to investigate VM-IAP heritability simultaneously allowed for the analysis of other factors that contribute to VM-IAP methylation states. When assessing whether the sex of an individual has an effect on its methylation level, a highly significant effect was observed at VM-IAP$_{Rnf157}$: males were more likely

to exhibit higher methylation levels than females (p = 1.20e-5; q = 2.16e-4; **Table 2.2**; **Figure 2.9**). Weaker evidence for a similar trend was observed for VM-IAP$_{Marveld2}$ and VM-IAP$_{Gm13849}$. This finding suggests that there may be sex-linked modifiers of VM-IAPs.

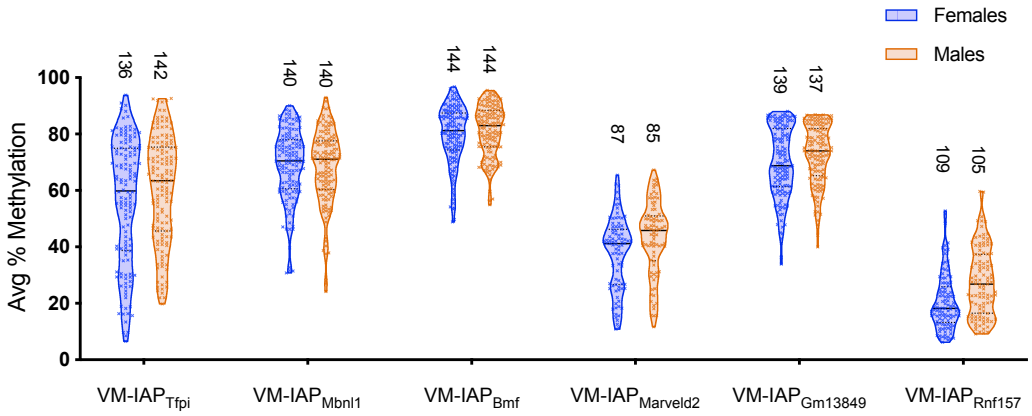


Figure 2.9: Methylation levels of F1 individuals used to fit LMMs, separated by sex. A significant sex effect was detected for VM-IAP$_{Rnf157}$ (p value: 1.20e-5; q-value: 2.16e-4; Table 2.2). Samples sizes are indicated above each violin plot. Violin plot elements: solid black line, median; dotted black lines, 25th and 75th percentiles; plot cut-offs, maximum and minimum; all data points shown inside violin plots.

2.3.3.3 Moving from ranges to frequency distributions

Having quantified the methylation level of six VM-IAPs in hundreds of offspring allowed for the re-evaluation of VM-IAP methylation dynamics in the B6 population. Previous data had shown that VM-IAP methylation state varies between two set boundaries in a locus-specific manner, but the frequency distribution within those boundaries was unknown. Frequency distribution histograms were generated for each VM-IAP using the average methylation value for each individual offspring (**Figure 2.10**). The resulting histograms were not uniform, revealing that all VM-IAP methylation values are not equally likely in the population. In addition, the data were not normally distributed: they displayed skewed bell curve shapes. This was tested and confirmed by standard normality tests and QQ plots (Appendix A, **Table A.3** and **Figure A.1**). These

frequency distributions provide locus-specific probabilities for each methylation value and will therefore be useful for future studies that aim to model the establishment and maintenance of VM-IAP methylation states. They will be revisited and discussed in more detail in Chapter 5 in conjunction with F1 hybrid datasets.

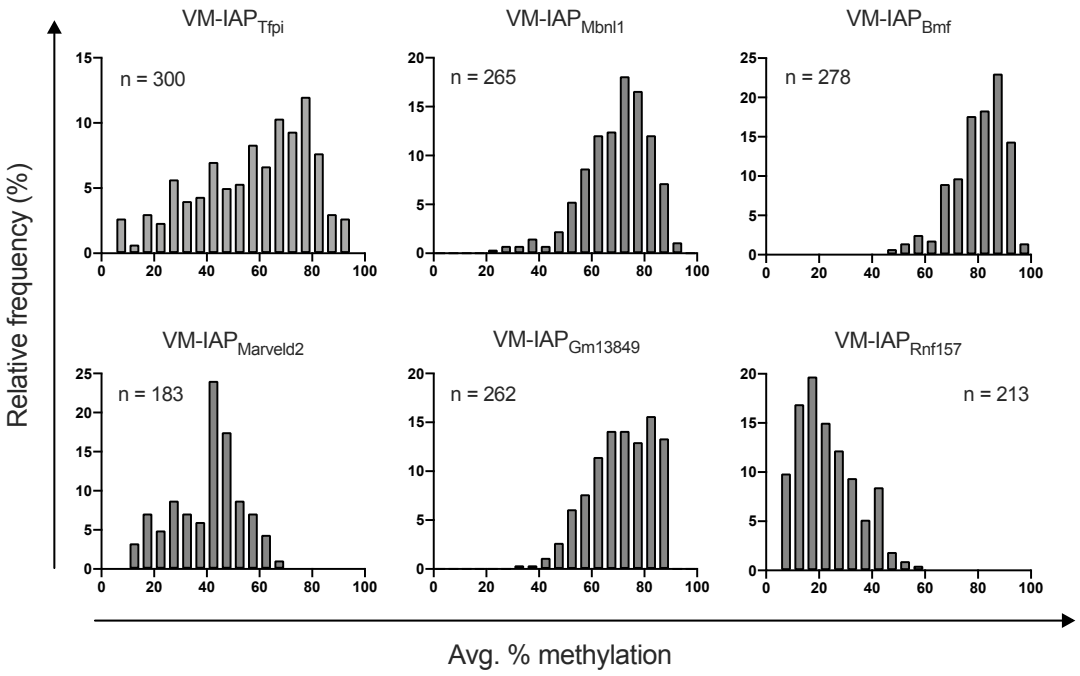


Figure 2.10: Frequency distribution histograms of VM-IAP methylation levels in the B6 population. Percent methylation values were averaged across CpGs at each locus and frequency distributions were generated by tabulating the relative frequency of values into 20 bins. Bin width was set to 5%. Sample sizes are shown on the graphs.

2.4 Discussion

Prior to the experiments presented in this chapter, the *A^vy* and *Axin^Fu* loci were the only metastable epialleles to have been investigated with regards to epigenetic inheritance (Morgan *et al.* 1999; Rakyan *et al.* 2003). These pivotal studies carried out in 1999 and 2003 injected optimism into the field of mammalian transgenerational epigenetic inheritance and suggested that other regions in the mouse genome, especially those with similar properties, may behave similarly. However, only one out of the six VM-IAPs examined here showed evidence of heritability through the maternal line reminiscent of *A^vy* inheritance dynamics. For the lone VM-IAP$_{Gm13849}$ locus, the effect size of the maternal contribution to offspring methylation level was distinctly small, raising questions about its biological relevance. Therefore, this work demonstrates that methylation variation at repeat elements is rarely associated with epigenetic inheritance and challenges the generalizability of this phenomenon at metastable epialleles.

Despite their lack of heritability, VM-IAPs conform with the *A^vy* and *Axin^Fu* metastable epialleles in that they reacquire a variable epigenetic state after passage through the germline. It is precisely this characteristic that led to the use of the term 'metastable' in the naming of metastable epialleles (Rakyan *et al.* 2002). This result squarely places VM-IAPs in this genomic category and expands our understanding of the inter-individual differences observed at metastable epialleles in the B6 genome. Studies on the *A^vy* and *Axin^Fu* loci have largely been based on qualitative phenotypic data collected by trained observers. In contrast, VM-IAPs were identified and characterised purely based on their methylation profiles, which enabled a more quantitative and comprehensive assessment of their epigenetic state. The results revealed that VM-IAP methylation distributions in the B6 population exhibit skewed bell-curve shapes that are highly locus specific. The frequency distributions of *A^vy* and *Axin^Fu* methylation levels have not yet been explored but are also expected to be non-uniform, since extreme phenotypes (e.g. fully yellow fur for *A^vy* mice or completely straight tails for *Axin^Fu* mice) are rarer than intermediate phenotypes in both mouse models.

Unlike the A^{vy} and $Axin^{Fu}$ loci, VM-IAPs are all found in the commonly used B6 inbred mouse strain. This facilitates comparative experiments across loci while keeping genetic background and experimental individuals constant. The co-variance analysis in this chapter took advantage of this and showed that the methylation level of a given VM-IAP is independent of methylation levels at other VM-IAPs within the same individual. This indicates that VM-IAP methylation levels are not determined by an individual-specific factor (e.g. varying expression levels of DNA methyltransferases in early development). This finding does not rule out a *trans*-acting mechanism that fosters metastability at repeat elements (which may be occurring, as discussed in Chapters 5 and 7), but it does demonstrate that the stochastic process that causes variable methylation levels is happening autonomously at each VM-IAP.

We suggest that it is the predictable reconstruction of epigenetic stochasticity from generation to generation that raises the most compelling mechanistic questions, rather than the subtle memory of parental methylation level observed at a minority of metastable epialleles. What is the role of genomic context in delimiting the consistent methylation distributions observed at each locus? What factors are at play during the probabilistic acquisition of methylation states within the limits of each distribution? At what developmental time point is parental methylation level forgotten, and when is the methylation state of offspring established? How long is the period of stochastic establishment and is a somatically heritable state established later in development? Which of these mechanistic aspects vary between VM-IAPs? The answers to these questions will provide considerable insight into the mechanisms that underlie the resetting of epigenetic variability across generations. This fundamental knowledge will subsequently provide the basis for understanding the partial epigenetic memory of parental methylation level observed at a small subset of loci.

Chapter 3

Developmental dynamics of VM-IAP methylation

3.1 Introduction and objectives

Chapter 2 documented the reconstruction of variable methylation at VM-IAPs across generations regardless of parental methylation level. This reconstruction of metastability should be considered in the context of the two rounds of global epigenetic reprogramming that occur in mammalian development: the first during preimplantation embryogenesis, and the second during early germ cell lineage specification. After fertilization, genome-wide DNA methylation erasure occurs via active and passive demethylation of the paternal and maternal genomes, respectively. The second genome-wide demethylation event occurs exclusively in primordial germ cells, after which sperm- and oocyte-specific methylation patterns are established (reviewed in Hackett & Surani 2013; Smallwood & Kelsey 2012). These processes are important for the reacquisition of cellular totipotency in the next generation, as well as the proper differentiation of the soma and the germ line.

For most regions in the genome, preimplantation epigenetic reprogramming is followed by the re-establishment of cell-type-specific epigenetic states that match those observed in the previous generation, resulting in high fidelity cell-type-specific methylomes (Hackett & Surani 2013). Global demethylation during

early development provides a mechanism by which parental VM-IAP methylation levels are erased, but does not explain the acquisition of stochastic methylation levels in the next generation. It has been reported that germline methylation levels of the A^{vy} and $Axin^{Fu}$ loci reflect methylation dynamics in the soma, and blastocysts were shown to be unmethylated at both regions regardless of parental methylation state (Appendix A, **Figure A.2**; Blewitt *et al.* 2006; Fernandez-Gonzalez *et al.* 2010; Rakyan *et al.* 2003). These findings led to the conclusion that these alleles are refractory to germline reprogramming but not to preimplantation reprogramming (Blewitt *et al.* 2006). It is unknown whether this is the case for VM-IAPs. The consistency in methylation levels across tissues within an individual exhibited by both VM-IAPs and classic metastable epialleles provides indirect evidence that their methylation state is established early in development, prior to germ layer differentiation, and perpetuated mitotically thereafter by maintenance methyltransferases (Kazachenka *et al.* 2018; Waterland *et al.* 2006; Waterland & Jirtle 2003).

The work described in this chapter aims to increase our understanding of the developmental dynamics of VM-IAP methylation and revisits some of the previously published data on the A^{vy} locus. As such, the chapter objectives are as follows:

1. To determine the methylation state of VM-IAPs in the male germline;

2. To replicate previous work on the A^{vy} locus in the male germline;

3. To establish the role of DNMT3C in the acquisition of methylation at VM-IAPs during spermatogenesis;

4. To investigate the time point in early development during which VM-IAP methylation states are established.

3.2 Results

3.2.1 Variable methylation is lost at VM-IAPs in the male germline.

Methylation levels at nine VM-IAPs were quantified in mature sperm from six to ten adult C57BL/6J (B6) males. Tail samples from the same individuals were used to assess their corresponding somatic tissue methylation levels. VM-IAP 5′ LTR methylation levels in tail DNA exhibited variable methylation across individuals. This was similar to the ranges and patterns observed in other somatic tissues from previous experiments (**Figure 3.1**; **Figure 2.2**). In contrast, all VM-IAPs were fully methylated in sperm in all individuals tested, and no evidence of variation was observed (**Figure 3.1**). Transposable elements analysed in bulk are heavily methylated in sperm, so this finding is in line with genome-wide dynamics (Kobayashi *et al.* 2012). However, it is inconsistent with previous studies on the A^{vy} and $Axin^{Fu}$ loci, where methylation levels at the IAP LTRs in sperm were reported to reflect the methylation state of somatic tissues despite the absence of paternal heritability of the coat colour phenotype (Blewitt *et al.* 2006; Rakyan *et al.* 2003).

The methylation levels of two imprinted loci, *Peg3* and *H19* (Kuroiwa *et al.* 1996; Tremblay *et al.* 1997), were assessed in bisulphite-converted sperm DNA samples. This was done to confirm that the observed methylation pattern in sperm was not due to failed bisulphite conversion, which would lead to apparent hypermethylation in all samples. In every case, paternally expressed *Peg3* was completely unmethylated while maternally expressed *H19* was fully methylated (**Figure 3.2**). This verified the success of the bisulphite conversion step and simultaneously acted as a sperm sample purity control, given that somatic cell contamination of male germ cells would result in a decrease in methylation at paternally methylated imprints and an increase in methylation at maternally methylated imprints.

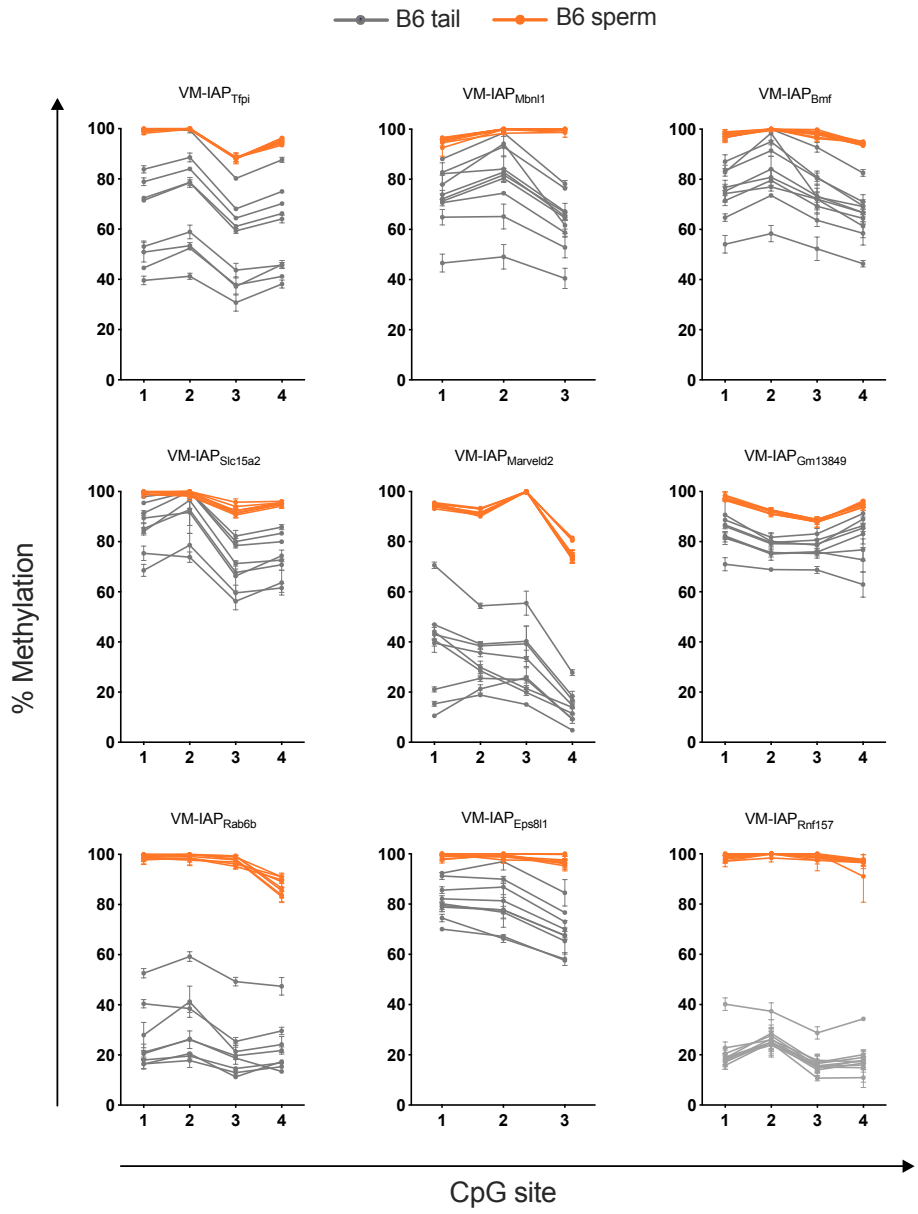


Figure 3.1: VM-IAPs are hypermethylated in mature sperm. DNA was extracted from sperm (orange) and tail (grey) samples collected from adult B6 males (n = 8-10). Percent methylation was quantified for the 3 to 4 most distal CpGs of the VM-IAP 5′ LTRs via bisulphite pyrosequencing. Each line represents data from one individual and error bars represent the standard deviation between technical triplicates.

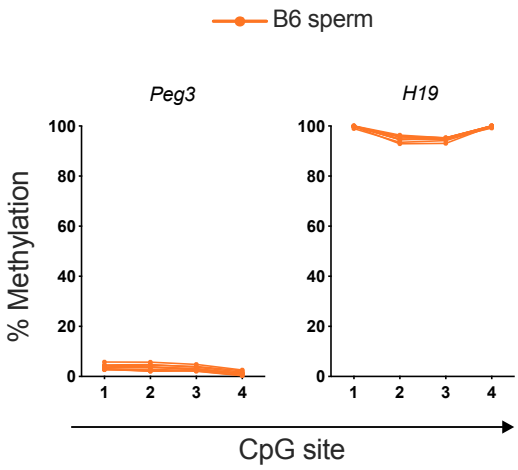


Figure 3.2: Assessment of sperm sample purity and bisulphite conversion efficiency. Sperm samples are unmethylated at paternally expressed *Peg3* and methylated at maternally expressed *H19* (n = 10). Percent methylation was quantified via bisulphite pyrosequencing. Each line represents data from one individual and each point represents the average of technical duplicates.

3.2.2 Revisiting the A^{vy} mouse model

The inconsistency between these VM-IAP results and previous studies on the A^{vy} and $Axin^{Fu}$ alleles in the male germline is puzzling. It is possible that there is an inherent biological difference between the classic metastable epialleles and the more recently identified VM-IAPs, as was the case with their heritability. The A^{vy} and $Axin^{Fu}$ insertional mutations arose in mouse strains different from the ones they are currently maintained on and hence constitute segments of DNA that are genetically foreign to the genome they reside in. This could potentially render them refractory to the sperm methylation machinery of the host genome. Alternatively, the contrasting findings between previous studies and the present one could reflect technical differences in the methods used to isolate sperm and/or quantify methylation.

To tease apart these two possibilities, we initiated a collaboration with the team of Dr. Greg Cox at The Jackson Laboratory (JAX), including his lab manager David Schroeder and research assistant Jennifer Stauffer. They cryo-recovered the A^{vy} mouse line maintained on a B6 background (B6.C3-Avy/J, JAX Stock No. 000017) and are currently collaborating with us on a number of experimental avenues. The first project we tackled aimed to replicate previously published

methylation data on A^{vy} sperm. It also served as a small-scale pilot experiment before initiating more breeding-intensive heritability studies, which are currently ongoing following the success of the experiments presented here.

The first experiment compared A^{vy} methylation levels between spleen and sperm samples collected from nine A^{vy}/a adult males that spanned the spectrum of coat colour phenotypes (including yellow, mottled, and pseudoagouti mice). A^{vy} methylation levels were quantified by bisulphite pyrosequencing using previously published primers that target the A^{vy} IAP 5′ LTR (Anderson *et al.* 2012). In an effort to avoid observer-expectancy effects and confirmation biases, the experiment was double-blinded: the coat colour quantification was carried out at JAX separately from the methylation analysis carried out in Cambridge, and the results were only compared after both analyses were complete. Coat colour was evaluated using two independent methods. The first method, performed by David Schroeder and reflective of previously published methods of A^{vy} coat colour quantification, consisted of visually classifying mice using a subjective coat colour score that ranges from 1 (yellow or nearly all yellow) to 4 (pseudoagouti or nearly all pseudoagouti). The second method was developed by Jennifer Stauffer based on a 2009 paper from the Cooney Lab (Ounpraseuth *et al.* 2009). This method determined coat colour proportions computationally based on photographic input and was therefore more objective and quantitative than the first. **Figure 3.3** shows photos of the nine male mice used for the initial study, along with their associated coat colour quantification using the two methods outlined above. Comparing the M8 mouse to the M9 mouse illustrates that the computational score is not necessarily more accurate than the visually determined one, highlighting the value of using two separate quantification methods. A more detailed comparison of the two methods can be found in Appendix A (**Figure A.3)**.

3.2.2.1 A^{vy} methylation levels are inversely correlated with coat colour phenotype

A^{vy} methylation levels quantified from spleen samples displayed wide-ranging inter-individual methylation variation, with the lowest value nearing

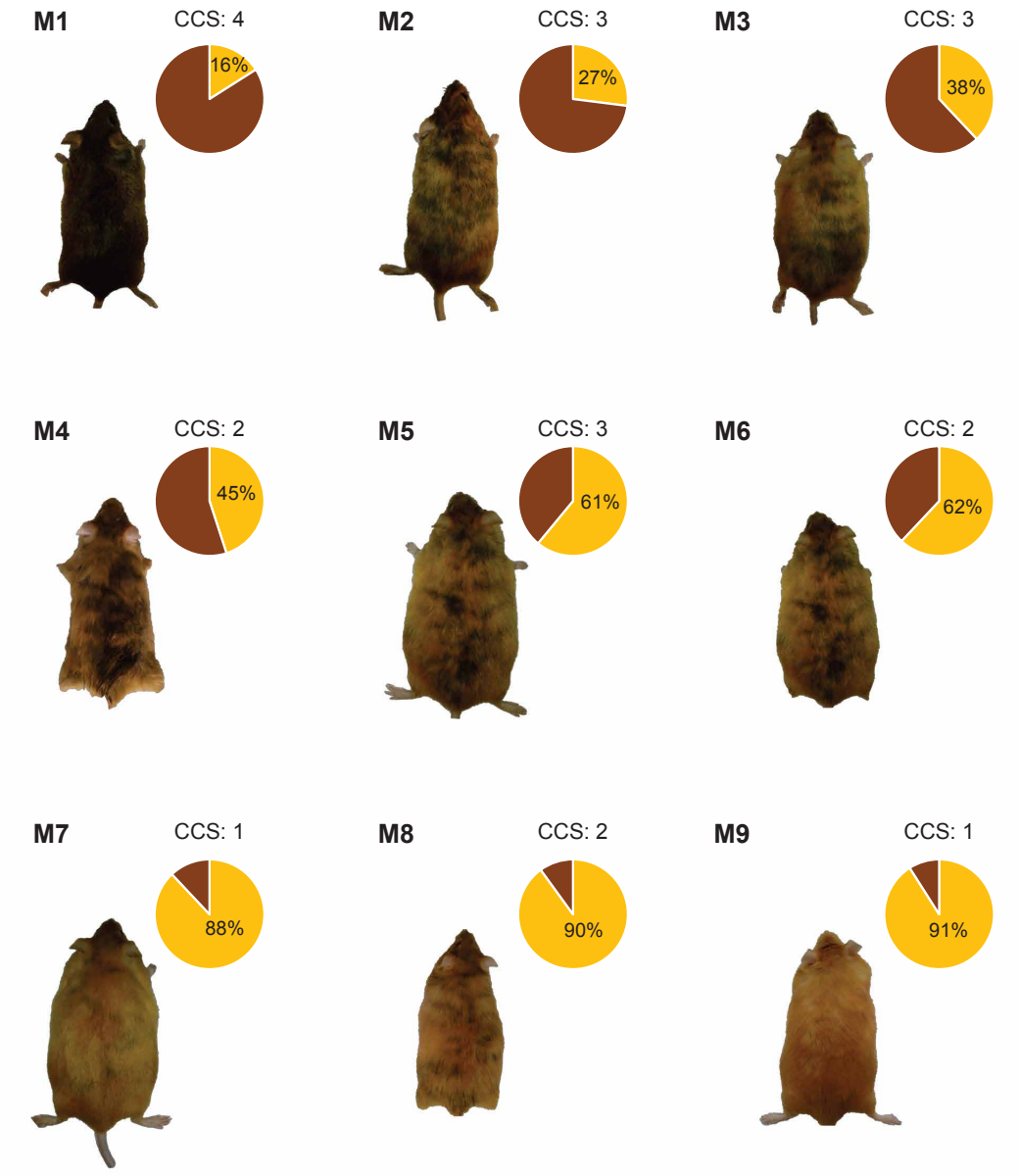


Figure 3.3: A^{vy}/a coat colour quantification. Nine A^{vy}/a males that span the A^{vy} coat colour spectrum were selected for a pilot experiment. Mice were visually classified by David Schroeder using a subjective coat colour score (CCS) that ranged from 1 (yellow or nearly all yellow) to 4 (pseudoagouti or nearly all pseudoagouti). Coat colour was also measured computationally by Jennifer Stauffer using photographic input. The output data is reported as percent of pelt area that is yellow, shown as pie charts (see also Appendix A, **Figure A.3**). Mice are labelled M1 through M9 and displayed in order of ascending percent yellow.

0% methylation and the highest exceeding 80% methylation (**Figure 3.4A**). This range resembled the one displayed by VM-IAP$_{Tfpi}$; both ranges are wider than the ranges observed for other VM-IAPs (**Figure 2.2**). A larger sample size of A^{vy}/a mice is required to evaluate the population frequency distribution of A^{vy} methylation levels. The correlation between A^{vy} methylation and coat colour was measured using the two coat colour quantification methods separately. Statistically significant inverse correlations were detected between methylation and coat colour using both methods (**Figure 3.4B**). This replicated published work and validated our methodology.

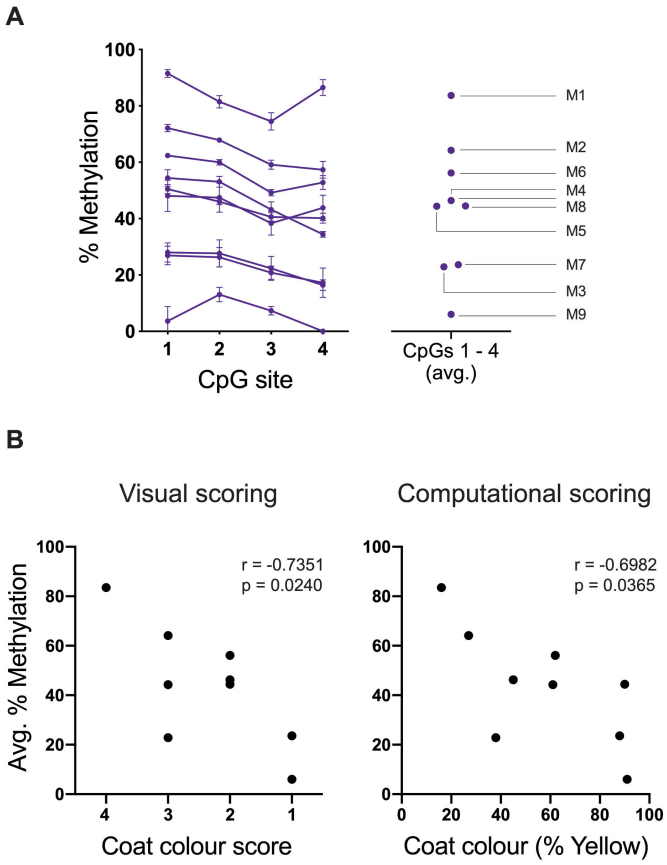


Figure 3.4: A^{vy} methylation levels are inversely correlated with coat colour. A. DNA methylation levels of the A^{vy} IAP 5′ LTR were measured in spleen samples collected from nine adult A^{vy}/a males spanning that span the phenotypic coat colour spectrum. Each line represents an individual and error bars represent the standard deviation between technical triplicates. The right-hand panel shows average methylation levels across CpGs for each individual, labelled to match the mouse designations used in **Figure 3.3**. **B.** Correlation analyses between average methylation and coat colour using two coat colour quantification methods. Pearson correlation coefficients (r) and p-values (p) are shown on the graphs.

3.2.2.2 *The A^{vy} allele is hypermethylated in sperm*

The methylation levels of imprinted loci *Peg3* and *IG-DMR* in A^{vy}/a sperm samples were consistent with the expected methylation levels in the male germline, verifying the success of the bisulphite conversion and the purity of the samples (**Figure 3.5**). Unlike A^{vy}/a spleen samples, all A^{vy}/a sperm samples exhibited high methylation levels at the A^{vy} IAP 5′ LTR irrespective of coat colour and somatic methylation level (**Figure 3.5**). This indicates that, at least in the B6 genetic background, the A^{vy} locus is targeted for methylation in the male germline. This result contrasts with previous findings in the literature (Blewitt *et al.* 2006; Rakyan *et al.* 2003) but is consistent with the hypermethylation of VM-IAPs in sperm.

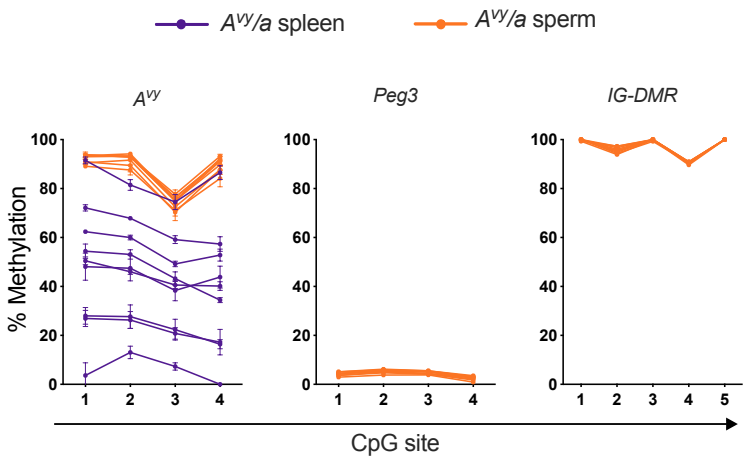


Figure 3.5: The A^{vy} locus is hypermethylated in the male germline. Methylation levels were quantified at the A^{vy} IAP 5′ LTR in mature sperm samples (orange) collected from the nine A^{vy}/a males shown in Figure 3.3. Spleen methylation levels in the sample individuals are shown for comparison (purple). Each line represents an individual and error bars represent the standard deviation between technical triplicates. The *Peg3* and *IG-DMR* imprinted loci were tested in sperm samples to assess purity and bisulphite conversion efficiency.

3.2.2.3 *VM-IAP methylation dynamics are conserved in A^{vy}/a mice*

The methylation states of six VM-IAPs were assessed in the same A^{vy}/a spleen and sperm samples used in the analysis above to determine whether their methylation profiles were conserved. All six VM-IAPs showed inter-individual methylation variation across A^{vy}/a spleens, with similar ranges to those observed in B6 tissues (**Figure 3.6**). Consistent with the results from section 3.2.1, VM-IAPs were hypermethylated and non-variable in A^{vy}/a sperm (**Figure 3.6**).

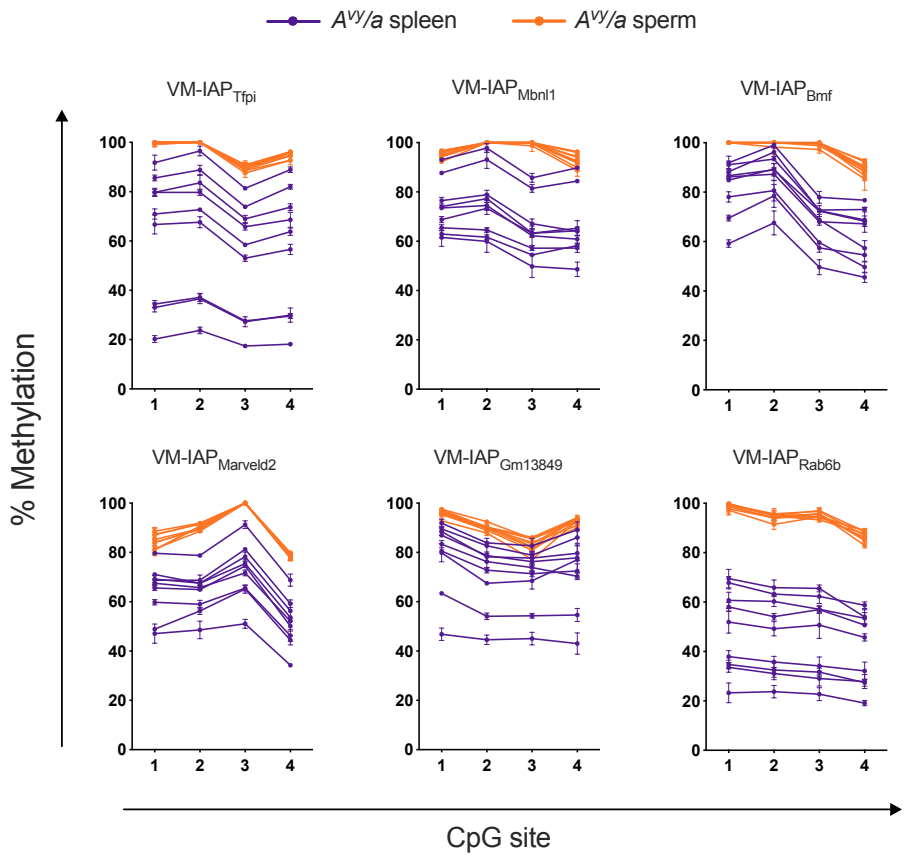


Figure 3.6: VM-IAP metastability is conserved in A^{vy}/a mice. Methylation levels at six VM-IAPs were quantified in A^{vy}/a sperm (orange) and spleen (purple) samples. Each line represents an individual and error bars represent the standard deviation between technical triplicates.

3.2.3 VM-IAPs are methylated in the male germline by DNMT3C

DNMT3C was recently discovered as a *de novo* DNA methyltransferase (Barau *et al.* 2016; Jain *et al.* 2017). The *Dnmt3C* gene is rodent-specific and evolved from a *Dnmt3B* duplication event. It is expressed exclusively in the testis and *Dnmt3C* knockout males fail to produce mature sperm (Barau *et al.* 2016). DNMT3C catalyses the methylation of young retrotransposons (mainly L1 promoters, but also IAPs) in the male germline (Barau *et al.* 2016). In light of the male germline-specific expression of *Dnmt3C* and the hypermethylation of VM-IAPs in sperm, it was hypothesized that DNMT3C is responsible for methylating VM-IAPs during spermatogenesis. This was tested in collaboration with Dr. Déborah Bourc'his and her former postdoctoral fellow Dr. Joan Barau. Having initially described this novel DNA methyltransferase, they provided us with germ cell and liver DNA collected from the *Dnmt3C* knockout mouse line. This mouse line harboured a deletion of *Dnmt3C* exon 15 and was generated using CRISPR-Cas9 gene editing technology in B6 mice (Barau *et al.* 2016).

3.2.3.1 Loss of VM-IAP methylation in Dnmt3C$^{KO/KO}$ germ cells

It is not possible to quantify methylation levels in *Dnmt3C$^{KO/KO}$* mature spermatozoa because spermatogenesis is halted at the pachytene stage of meiosis I in *Dnmt3C$^{KO/KO}$* males. Instead, germ cells were collected by Dr. Joan Barau from *Dnmt3C$^{WT/WT}$*, *Dnmt3C$^{KO/WT}$*, and *Dnmt3C$^{KO/KO}$* males at postnatal day 10 (P10). The germ cells were sorted by fluorescence-activated cell sorting (FACS) with the EpCAM antibody, a marker for progenitor spermatogonia (Kanatsu-Shinohara *et al.* 2011). VM-IAP methylation levels were quantified in the P10 germ cell samples to assess whether *Dnmt3C* is responsible for the methylation of VM-IAPs in the male germline.

Dnmt3C$^{WT/WT}$ P10 germ cells were highly methylated at the 11 VM-IAPs tested, reflecting previous findings in mature sperm. This finding indicates that VM-IAPs acquire methylation in the male germ line before differentiating into spermatozoa (**Figure 3.7**). VM-IAP methylation levels in heterozygote *Dnmt3C$^{KO/WT}$* P10 germ cells reflected the levels observed in wild type germ cells, which is consistent with previous molecular and phenotypic analyses on this

mouse line (**Figure 3.7**; Barau *et al.* 2016). *Dnmt3C*$^{KO/KO}$ germ cells showed a severe loss in methylation at VM-IAPs, indicative of DNMT3C-mediated methylation (**Figure 3.7**; **Figure 3.8**). A bulk assay that targeted the LTR of more than 1000 highly conserved IAPEz elements also showed a significant decrease in methylation levels in *Dnmt3C*$^{KO/KO}$ germ cells compared to *Dnmt3C*$^{WT/WT}$ and *Dnmt3C*$^{KO/WT}$ germ cells. This is in line with a previous experiment conducted by Barau and colleagues using the same primer set (**Figure 3.7**; Barau *et al.* 2016). However, the decrease in methylation was much smaller than that observed for VM-IAPs, suggesting that the methylation state of a minority of IAPs in the B6 genome is affected by the *Dnmt3C* knockout. This is fortified by our analysis of two individual non-variable IAP elements, IAP$_{Zak}$ and IAP$_{Cdk15}$, which both showed only modest reductions in methylation in *Dnmt3C*$^{KO/KO}$ samples (**Figure 3.7**).

As in our previous experiments, imprinted loci were used as controls for methylation analysis of the P10 germ cells. Maternally expressed *H19* was highly methylated and paternally expressed *Peg3* was lowly methylated in all samples irrespective of *Dnmt3C* genotype (**Figure 3.7**). However, *H19* methylation levels were slightly lower and *Peg3* methylation levels slightly higher than anticipated, and the variability between individuals was greater than that observed in previous experiments using mature sperm samples (**Figure 3.7**). Because DNA methylation at imprinted DMRs are established prior to the spermatogonia stage (Sasaki & Matsui 2008), this observation is likely due to the presence of contaminating somatic cells that were non-specifically bound by the antibody during FACS. This is a common difficulty with this type of experiment and may explain why the *Dnmt3C*$^{WT/WT}$ and *Dnmt3C*$^{KO/WT}$ germ cells are not 100% methylated, as well as why sample-specific methylation patterns are observed (Figure **3.8**). In addition, this implies that *Dnmt3C*$^{KO/KO}$ germ cells are more lowly methylated than is indicated on the graphs in **Figure 3.7**.

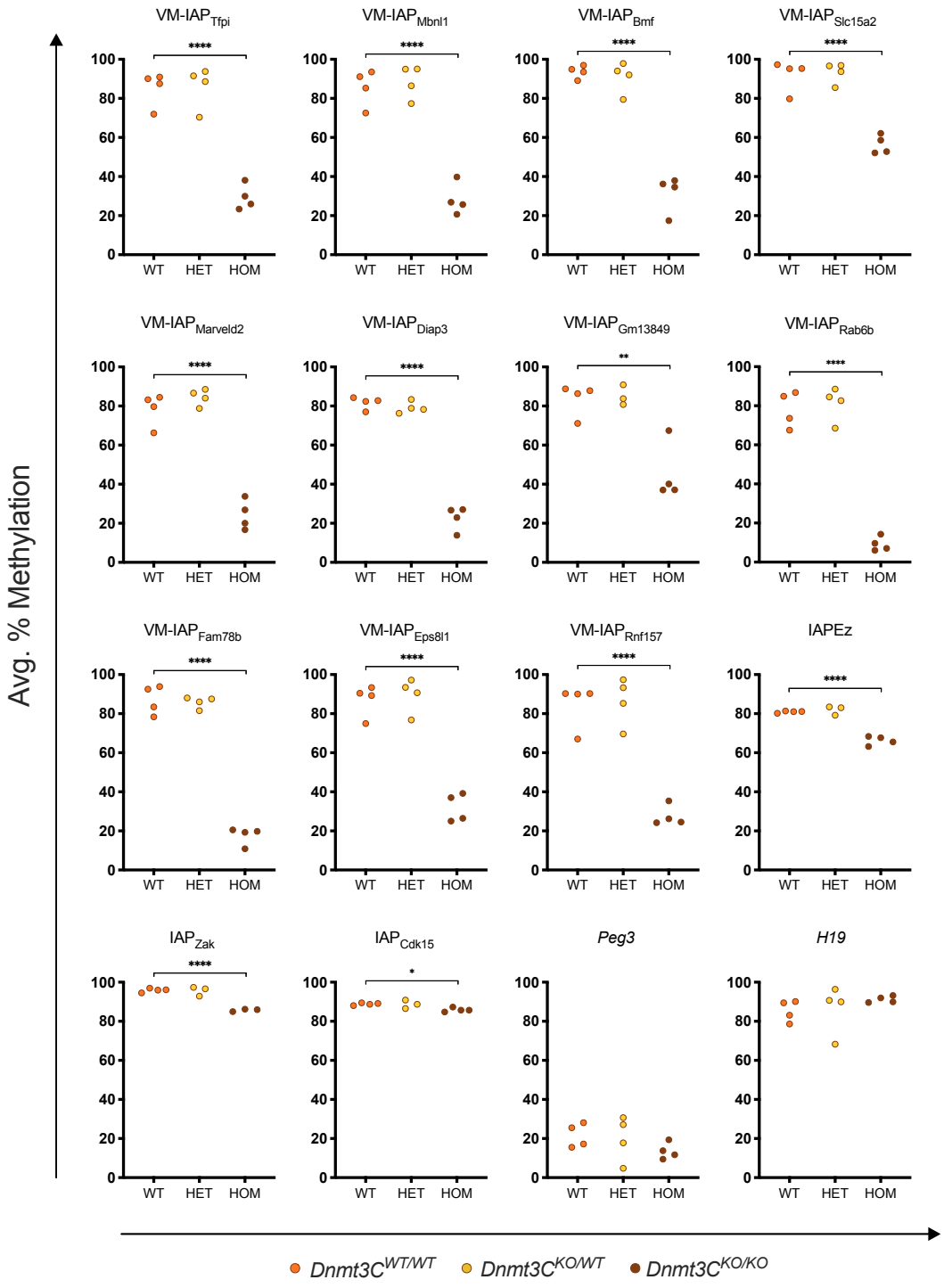


Figure 3.7: DNMT3C methylates VM-IAPs in the male germline. Methylation levels were quantified in *Dnmt3C*$^{WT/WT}$ (WT; orange circles), *Dnmt3C*$^{KO/WT}$ (HET; yellow circles), and *Dnmt3C*$^{KO/KO}$ (HOM; brown circles) male germ cells isolated by FACS from P10 male pups. Interrogated loci included 11 VM-IAPs, IAP$_{Zak}$, IAP$_{Cdk15}$, conserved IAPEz elements, and imprinted loci *Peg3* and *H19*. Plotted values represent the average methylation across CpGs. Each data point represents an individual (n = 4 per genotype). Genotypes were compared by one-way ANOVA using Dunnett's multiple comparisons test (* $p < 0.05$; ** $p < 0.005$; **** $p < 0.0001$).

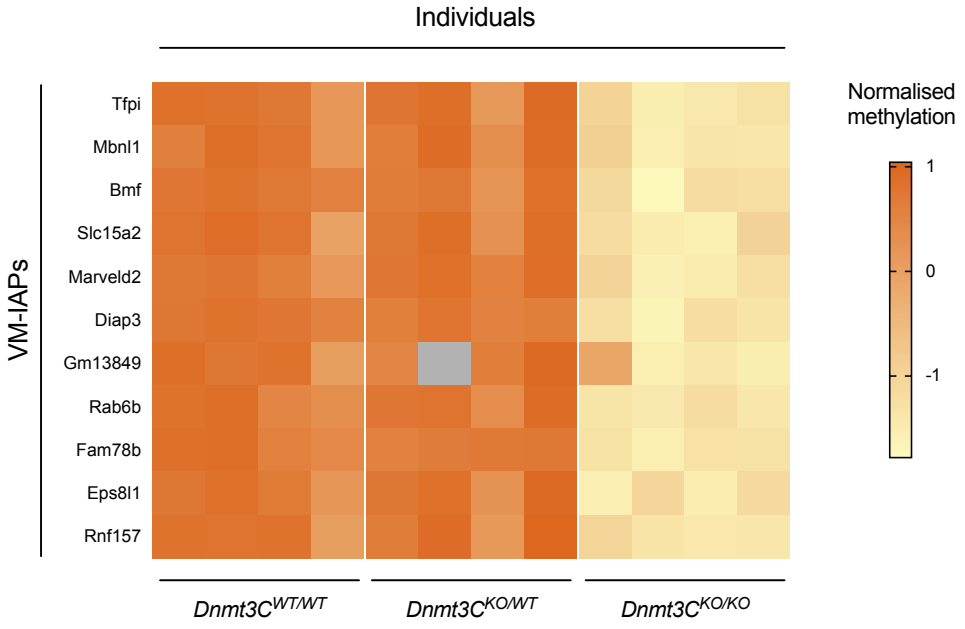


Figure 3.8: Individual-specific methylation patterns in FACS-isolated P10 germ cells. Heatmap compares VM-IAP methylation levels in *Dnmt3C*$^{WT/WT}$, *Dnmt3C*$^{KO/WT}$, and *Dnmt3C*$^{KO/KO}$ germ cells (n = 4 per genotype). Values were normalised to the inter-individual methylation range of each locus. Rows represent VM-IAPs and columns represent individual samples, separated by genotype. For technical reasons, VM-IAP$_{Gm13849}$ methylation levels could not be determined for one of the *Dnmt3C*$^{KO/WT}$ samples (grey square).

3.2.3.2 DNMT3C may play a role VM-IAP methylation in the soma

To verify that somatic tissues of *Dnmt3C*$^{WT/WT}$, *Dnmt3C*$^{KO/WT}$, and *Dnmt3C*$^{KO/KO}$ individuals retained metastability at VM-IAPs, the same methylation analysis previously conducted on DNA from P10 germ cells was carried out on liver DNA from the same individuals. As expected, methylation levels were high and invariable at control IAPs and intermediate at imprinted loci (**Figure 3.9A**). VM-IAPs exhibited inter-individual methylation variation, indicating that their characteristic epigenetic feature is conserved in the *Dnmt3C* knockout mouse line (**Figure 3.9B**). However, upon closer examination of the data, VM-IAP methylation levels appear to cluster by *Dnmt3C* genotype (**Figure 3.9B**). Unlike the germ cell data which indicated equivalent methylation levels

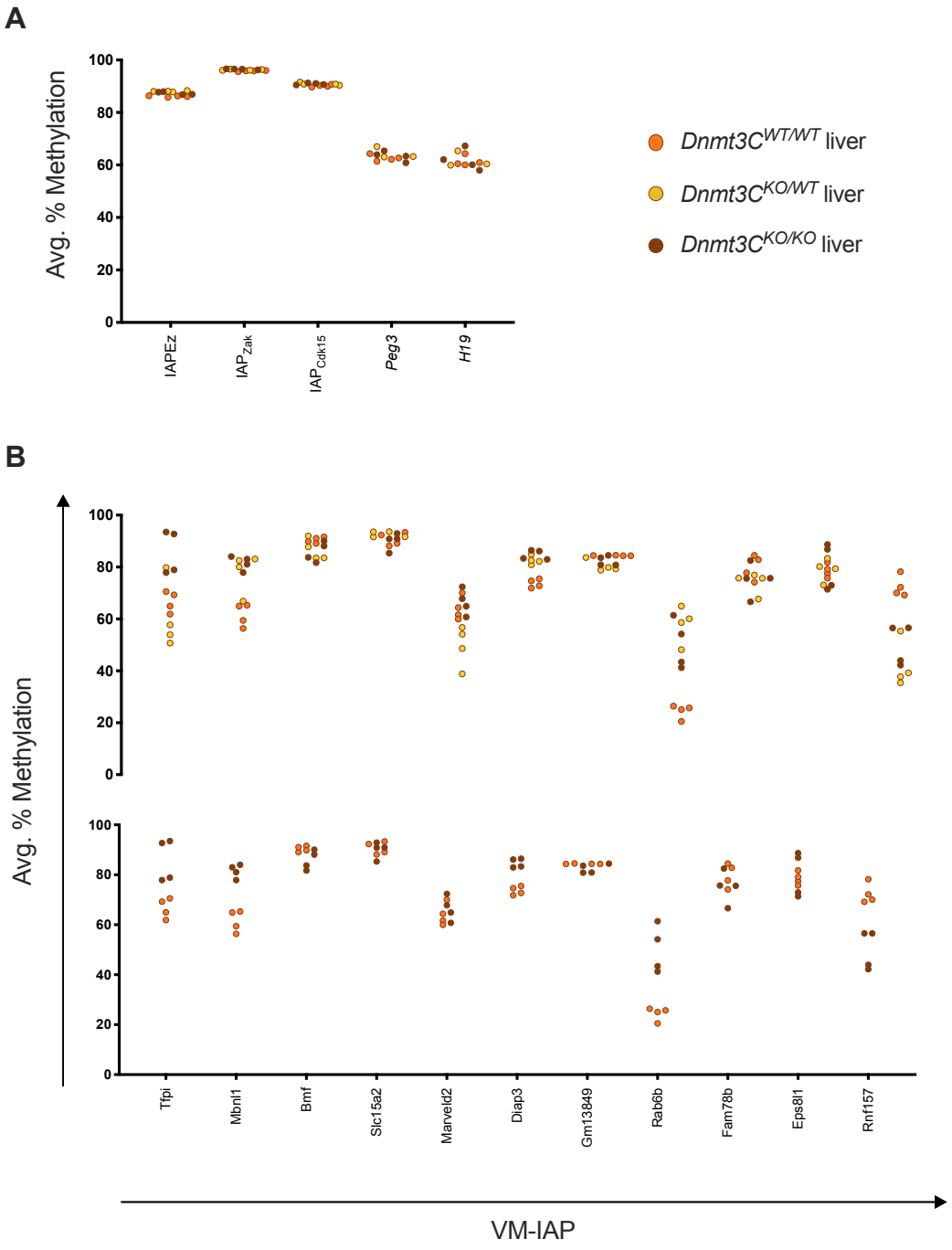


Figure 3.9: DNMT3C may regulate VM-IAPs in the soma. Methylation levels were quantified in *Dnmt3C*$^{WT/WT}$ (orange circles), *Dnmt3C*$^{KO/WT}$ (yellow circles), and *Dnmt3C*$^{KO/KO}$ (brown circles) liver tissue collected from the same P10 pups used for the germ cell analysis. **A.** Methylation data for conserved IAPEz elements, IAP$_{Zak}$, IAP$_{Cdk15}$, and control imprinted loci *Peg3* and *H19*. **B.** Methylation data for 11 VM-IAP 5′ LTRs. The *Dnmt3C*$^{KO/WT}$ samples have been removed from the lower graph to improve visualization of data clustering. Plotted values represent the average methylation across CpGs. Each data point represents an individual (n = 4 per genotype).

between *Dnmt3C*$^{WT/WT}$ and *Dnmt3C*$^{KO/WT}$ mice, the liver data showed clustering of *Dnmt3C*$^{KO/WT}$ and *Dnmt3C*$^{KO/KO}$ mice, albeit for only a subset of VM-IAPs. *Dnmt3C*$^{KO/WT}$ and *Dnmt3C*$^{KO/KO}$ mice were generated by crossing *Dnmt3C*$^{KO/KO}$ females to *Dnmt3C*$^{KO/WT}$ males, while *Dnmt3C*$^{WT/WT}$ mice were generated by *Dnmt3C*$^{KO/WT}$ intercrosses. The clustering may therefore be due to maternal depletion of *Dnmt3C*. *Dnmt3C*$^{KO/WT}$ and *Dnmt3C*$^{KO/KO}$ individuals were often more highly methylated than *Dnmt3C*$^{WT/WT}$ individuals, potentially pointing to a dominant negative effect over other DNMTs in early development (**Figure 3.9B**). This analysis needs larger sample sizes of each genotype before conducting statistical analyses or drawing firm conclusions. Nonetheless, this preliminary result argues that DNMT3C activity influences the establishment of VM-IAP methylation in the soma, which contrasts with previous reports indicating that *Dnmt3C* expression is restricted to the male germline (Barau *et al.* 2016). DNMT3C may therefore play a currently unappreciated role in somatic epigenetic regulation.

3.2.4 VM-IAP methylation states are likely established during preimplantation development

The first cell lineage specification in early mouse development separates the inner cell mass (ICM) from the trophectoderm (TE). The ICM eventually gives rise to embryonic tissues and the TE generates extraembryonic tissues, including the placenta. These two cell lineages become distinct from one another at the 32-cell stage, at which point the embryo undergoes cavitation to become a blastocyst, characterised by the formation of the blastocoel cavity containing the ICM and surrounded by an outer trophoblast layer. At this point the embryo has not yet implanted into the uterus, hence why the 4.5 days following fertilisation are referred to as preimplantation development (reviewed in Jedrusik 2015; Marikawa & Alarcón 2009).

3.2.4.1 VM-IAPs are hypomethylated but retain variability in E16.5 placentas

The analysis of VM-IAP methylation states in the placenta provides insight into the developmental time point of their establishment. Embryonic tails and

whole placentas were dissected at E16.5. All tested VM-IAPs showed significantly lower methylation levels in placenta compared to corresponding tail samples (**Figure 3.10**; paired t-tests). This is consistent with genome-wide

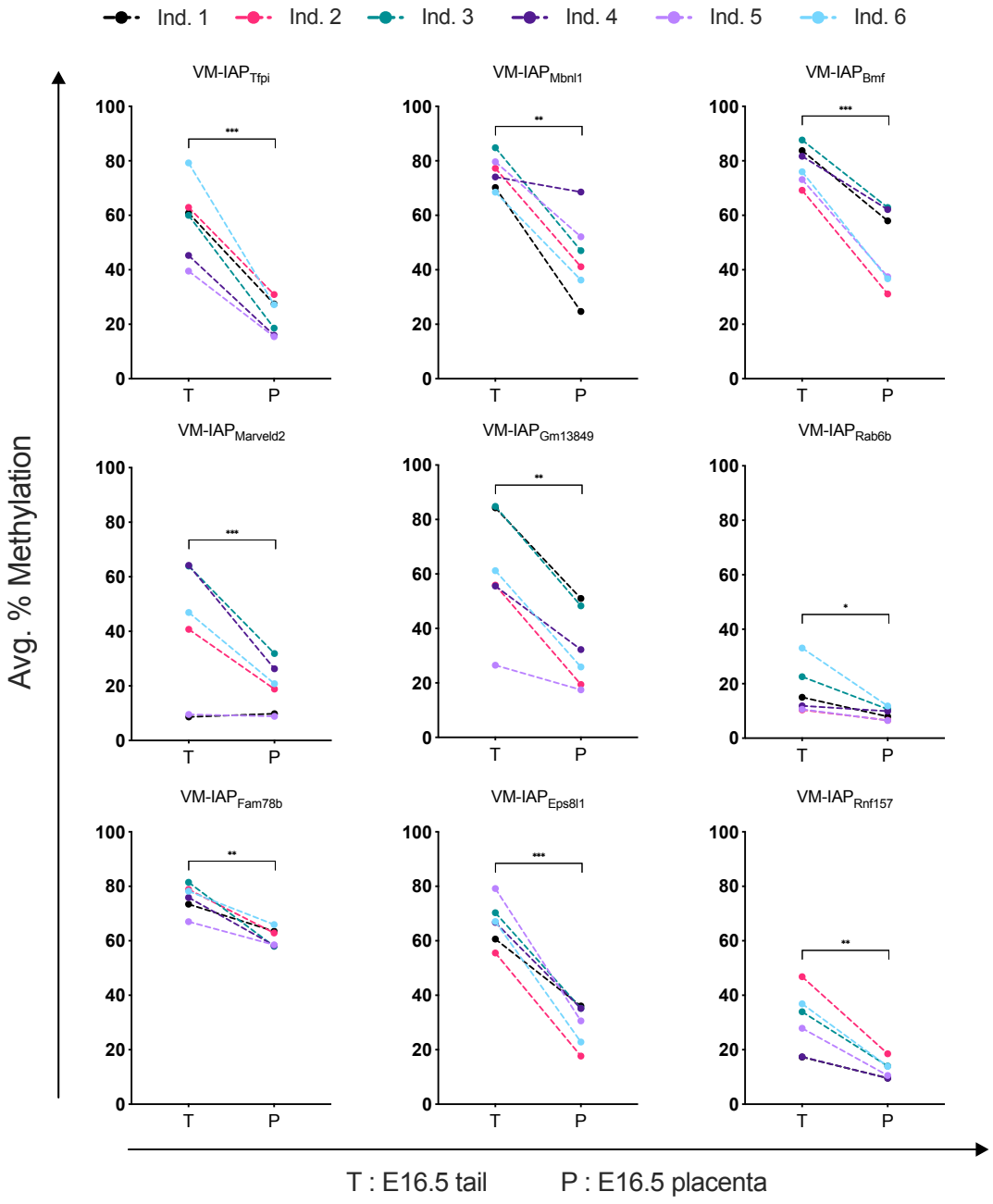


Figure 3.10: VM-IAPs are hypomethylated and variable in E16.5 placentas. VM-IAP 5′ LTR methylation levels were quantified in matched E16.5 tail and placenta samples (n = 6). Matched samples are colour-coded and connected by a dashed line. Plotted values represent the average methylation across CpGs. Tail and placenta methylation levels were compared via paired t-tests (* $p < 0.05$; ** $p < 0.005$; *** $p < 0.0005$).

hypomethylation observed in mammalian placental methylomes compared to embryonic methylomes (Ehrlich *et al.* 1982; Schroeder *et al.* 2015). Despite the reduced methylation levels, inter-individual methylation variation was retained in placental tissues for most VM-IAPs (**Figure 3.10**). In addition, statistically significant correlations were observed between placental and

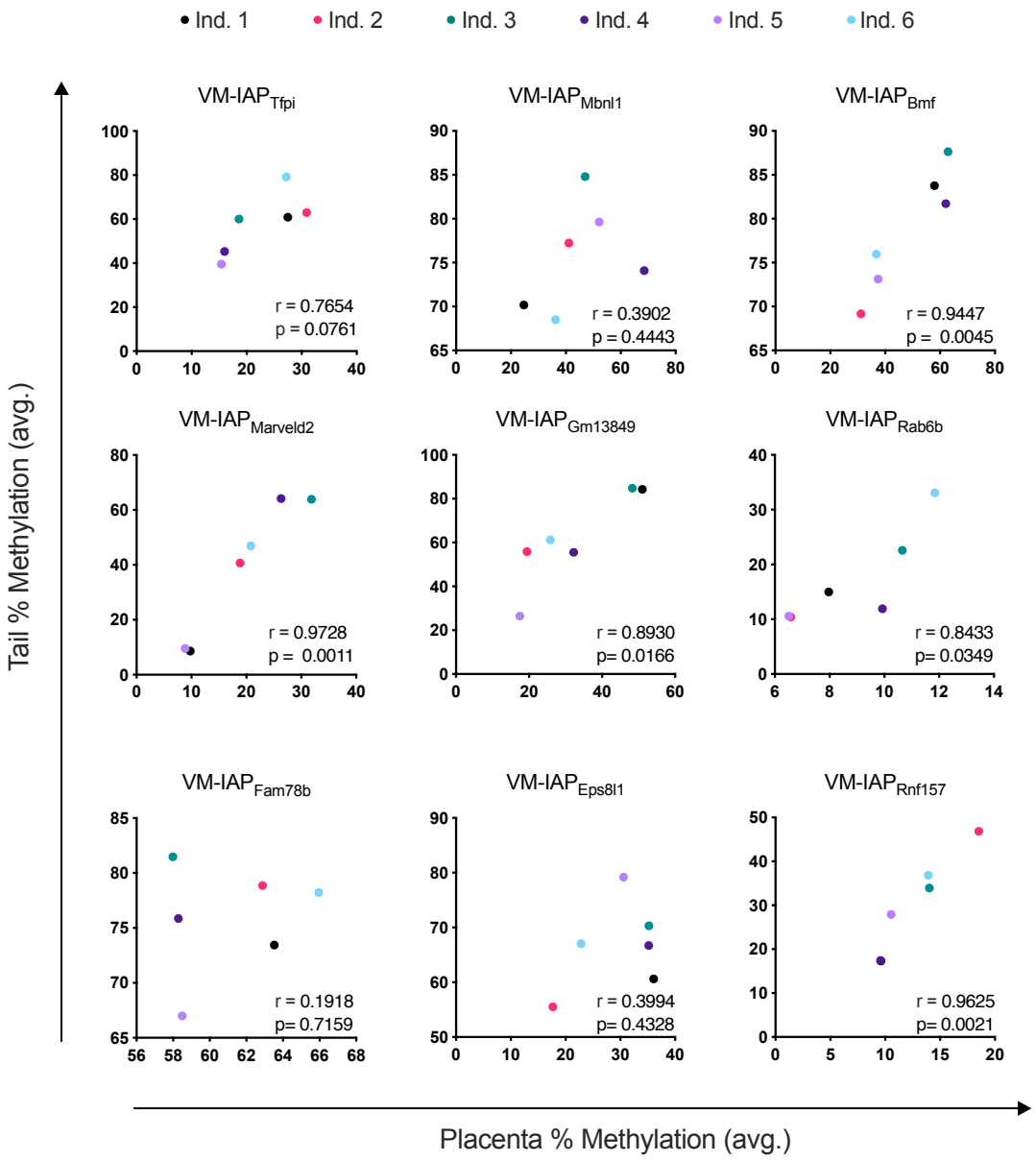


Figure 3.11: Tail and placental methylation levels are correlated for most VM-IAPs. Individuals are colour-coded as in **Figure 3.10**. Plotted values represent the average methylation across CpGs. Pearson correlation coefficients (r) and p-values (p) are shown on each graph.

corresponding tail methylation levels for most VM-IAPs (**Figure 3.11**). This indicates that the stochastic process that governs the establishment of methylation levels at VM-IAPs occurs (at least partially) prior to the specification of the ICM and TE. A set of eight IAPs that are normally highly methylated and non-variable in somatic tissues (genomic coordinates in Appendix A, **Table A.2**) also showed significantly lower methylation levels in placenta, but most remained non-variable and above 70% methylation (**Figure 3.12**). IAP_{Cdc73} and $IAP_{Slc124a3}$ exhibited the most substantial decreases, with $IAP_{Slc24a3}$ showing narrow-ranging inter-individual variation suggestive of a possible placenta-specific VM-IAP (**Figure 3.12**). This is an interesting possibility given that Jessica Elmer in our group has recently identified tissue-specific VM-IAPs that only exhibit variable methylation in certain cell types.

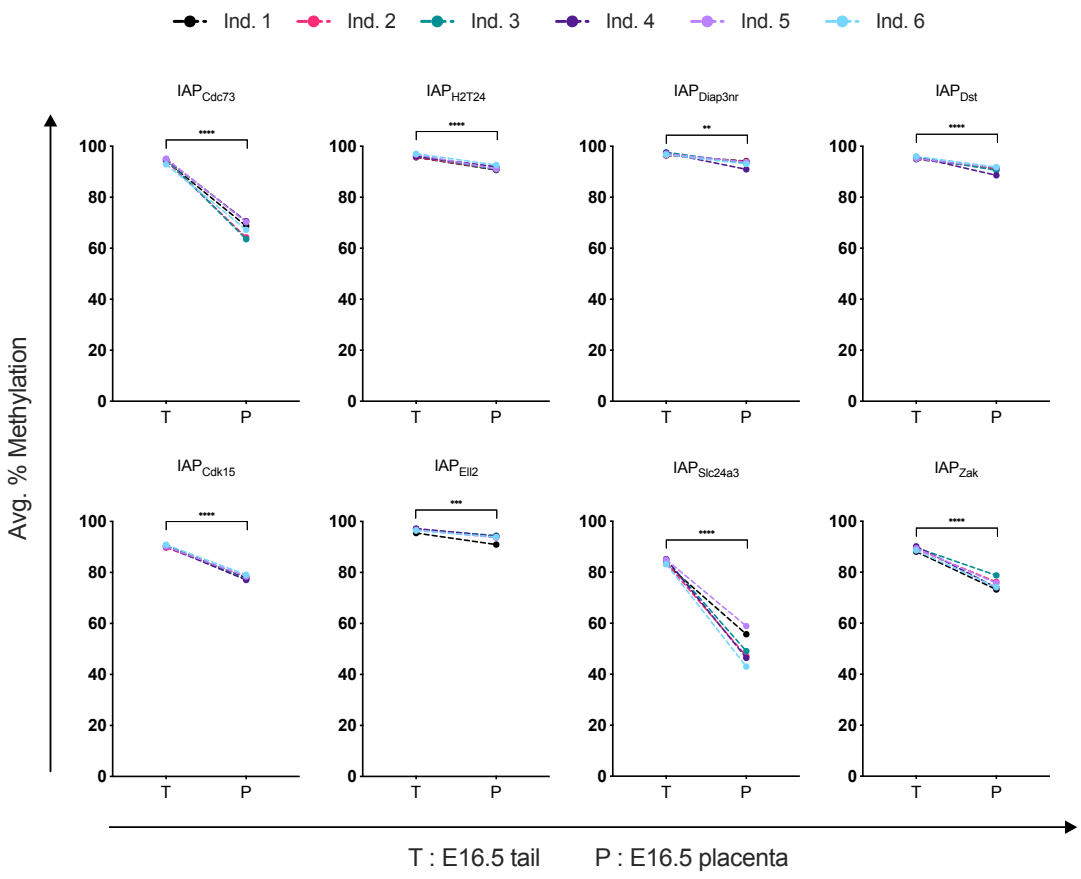


Figure 3.12: Non-variable IAPs are moderately hypomethylated in E16.5 placentas. IAP 5′ LTR methylation levels were quantified in matched E16.5 tail and placenta samples (n = 6). Matched samples are colour-coded and connected by a dashed line. Plotted values represent the average methylation across CpGs. Tail and placenta methylation levels were compared via paired t tests (** $p < 0.005$; *** $p < 0.0005$; **** $p < 0.0001$).

3.2.4.2 VM-IAPs exhibit variable methylation in embryonic stem cells

It is difficult to obtain high-quality quantitative methylation data when working with low DNA concentrations. This issue is pertinent when assessing methylation levels in individual preimplantation embryos: technical replicates deviate from one another due to the random preferential amplification of methylated or unmethylated target copies, which causes the results to be unrepresentative of the methylated-to-unmethylated ratio in the original sample. This problem is inconsequential when the locus in question is fully methylated or unmethylated, but becomes a significant hindrance when working with intermediately methylated regions such as VM-IAPs. A common solution to this issue is the pooling of samples to bolster the concentration of the starting material, but this is not ideal for the purposes of the present study because the focus is on inter-individual differences in methylation.

To avoid this technical difficulty and still gain insight into VM-IAP methylation dynamics in the early embryo, VM-IAP methylation was investigated in embryonic stem cells (ESCs). Four ESC lines were generated from the ICM of B6 male blastocysts and genomic DNA was extracted two or three passages after the establishment of each line. The detailed protocol used to generate the ESCs can be found in Chapter 8. For the majority of VM-IAPs, variable methylation levels were observed between the four cell lines (**Figure 3.13**). In a side-by-side comparison to ear methylation profiles, ESCs appeared to lie within the normal range observed at each VM-IAP in the B6 population (**Figure 3.13**). VM-IAP$_{Fam78b}$ was the exception, showing much lower methylation levels in ESCs compared to somatic methylation levels (**Figure 3.13**). The methylation levels of the imprinted loci *Peg3* and *H19* were around 50% methylation and none of the cell lines exhibited consistently low or high methylation levels across the VM-IAPs. This suggests that the variable nature of VM-IAPs in ESCs in not due to a disruption in global methylation patterns related to the culture conditions of the cells. Under the contentious assumption that ESCs are a reflection of the ICM, these results suggest that the establishment of stochastic methylation at VM-IAPs occurs prior to implantation.

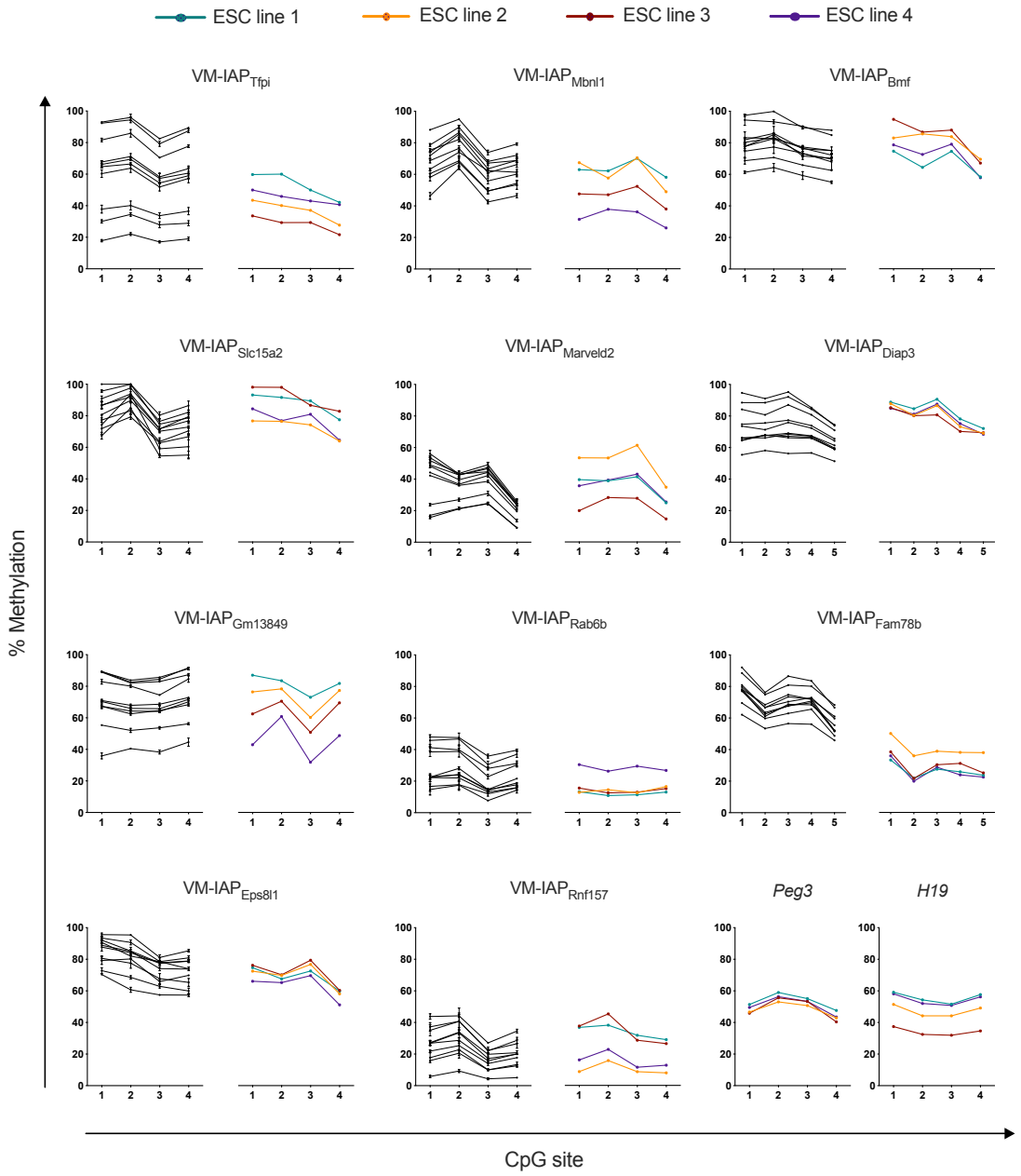


Figure 3.13: VM-IAPs exhibit variable methylation across ESC lines. Methylation levels were quantified in four ICM-derived male ESC lines. Tested loci included 11 VM-IAP 5′ LTRs and control imprinted genes *Peg3* and *H19*. Each line represents data from one cell line (colour-coded) and each point represents the average of technical duplicates. ESC data are presented next to the ear methylation data from **Figure 2.2** (black lines) to facilitate comparison of inter-individual ranges.

3.2.4.3 *VM-IAP$_{Fam78b}$ blastocyst methylation levels mirror ESC and oocyte methylation profiles*

ICM-derived ESCs can be of great value as an *in vitro* model to study preimplantation dynamics, but must be used with caution because culture conditions and passage number can both have profound effects on the DNA methylation landscape of ESCs (Lee *et al.* 2018; unpublished observations by members of the Ferguson-Smith Lab). VM-IAP$_{Fam78b}$ was the only VM-IAP to exhibit methylation levels in ESCs that were far below the levels normally observed in postnatal somatic tissues. To test whether this result was a technical artefact, *in vivo* VM-IAP$_{Fam78b}$ methylation levels were assessed in blastocysts. E3.5 blastocysts were dissected from three pregnant B6 females and embryo littermates (5-8) were pooled before bisulphite conversion and pyrosequencing. All three pooled samples exhibited methylation levels comparable to those observed in ESC lines, suggesting that ESCs are an appropriate model to study VM-IAP methylation in the preimplantation embryo (**Figure 3.14**).

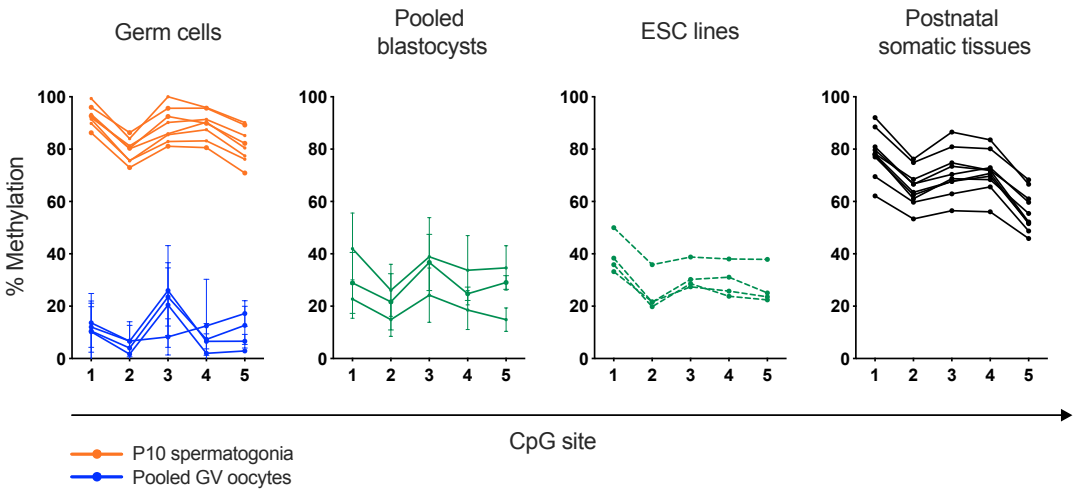


Figure 3.14: VM-IAP$_{Fam78b}$ methylation dynamics during development. Methylation levels at the VM-IAP$_{Fam78b}$ 5′ LTR were quantified in four pools of germinal vesicle (GV) oocytes (blue lines, 100 oocytes per pool) and three pools of blastocysts (green lines, 5-8 blastocysts per pool). Error bars represent the standard deviation between technical triplicates. P10 male germ cell data (orange lines) were taken from **Figure 3.7**. ESC data (dashed green lines) were taken from **Figure 3.13**. Postnatal ear data (black lines) were taken from **Figure 2.2**. Oocytes and blastocysts were quality checked by Dr. Nozomi Takahashi, an expert on germ cell and early embryo morphology.

These *in vitro* and *in vivo* results suggest that VM-IAP$_{Fam78b}$ acquires its final methylation state at a later developmental time point compared to other VM-IAPs. In addition, the low methylation levels indicate that the paternal allele, highly methylated in sperm (**Figure 3.1**), is fully demethylated following fertilisation. To further understand the developmental dynamics of this VM-IAP, methylation levels were quantified in pooled germinal vesicle (GV) oocytes collected from B6 ovaries (four pools of 100 oocytes each). These were completely unmethylated (**Figure 3.14**). It is possible that the lowly methylated maternal allele contributes to the unique behaviour of VM-IAP$_{Fam78b}$ compared to the other VM-IAPs, but additional loci must be examined in the female germline to test this relationship.

3.3 Discussion

The results in this chapter demonstrate that VM-IAPs are methylated in the male germline before the reconstruction of their variable state after fertilization, indicative of reprogramming upon paternal inheritance. This result contrasted with previously published findings that reported that the methylation state of the A^{vy} and $Axin^{Fu}$ alleles in sperm reflect the methylation state in somatic tissues of the same individual (Appendix A, **Figure A.2**; Blewitt *et al.* 2006; Rakyan *et al.* 2003). An attempt to replicate this A^{vy} result revealed that the A^{vy} IAP is heavily methylated in sperm and reprogrammed after fertilisation. PCR amplification bias or somatic cell contamination are potential explanations for these conflicting results, but the original A^{vy}/a sperm samples would have to be revisited to resolve the discrepancy. A^{vy} hypermethylation in the male germline might explain the lack of heritability observed upon paternal transmission in a B6 genetic background (Morgan *et al.* 1999). Since paternal inheritance does occur at the $Axin^{Fu}$ locus, $Axin^{Fu}$ methylation in sperm may in fact reflect somatic methylation patterns (Rakyan *et al.* 2003). Rakyan and colleagues showed that inheritance via the paternal lineage in $Axin^{Fu}$/+ mice is linked to the 129/Sv genetic background, so it is possible that the efficiency of epigenetic targeting of variably methylated repeat elements during spermatogenesis varies across mouse strains.

As reviewed in Chapter 1, it has been reported that IAP elements are resistant to global demethylation in the germline and during preimplantation development (Lane *et al.* 2003; Seisenberger *et al.* 2012). Although this initially provided a potential mechanism by which the IAP-driven A^{vy} and $Axin^{Fu}$ phenotypes are inherited, published work has found that both loci are demethylated at the blastocyst stage regardless of parental phenotype. This indicates that DNA methylation is not the mark that mediates phenotypic heritability (Blewitt *et al.* 2006; Fernandez-Gonzalez *et al.* 2010). Coupled with the finding that distinct VM-IAP methylation levels are not inherited (Chapter 2), this suggests that metastable epialleles are more susceptible to postfertilisation reprogramming than their non-variable IAP counterparts. Perhaps the inability of metastable epialleles to remain methylated throughout reprogramming allows for the acquisition of variable methylation states. The

Dnmt3C knockout experiments described in section 3.2.3.1 highlight this property. VM-IAPs were unmethylated in *Dnmt3C*$^{KO/KO}$ germ cells while non-variable IAPs showed a much weaker loss in methylation. Rather than reflecting VM-IAP-specific DNMT3C targeting, this finding is more likely the result of differential resistance to the erasure of methyl marks in the early stages of male germ cell development. This is supported by the observation that VM-IAP methylation levels in *Dnmt3C*$^{KO/KO}$ germ cells were lower than the levels normally observed at each locus in somatic tissues.

The significant correlations observed between embryonic tail and placental samples for most VM-IAPs suggest that the mechanisms that lead to stochastic methylation are active prior to the differentiation of the trophectoderm. This may involve the establishment of a baseline methylation level in the early embryo that is subsequently and separately modified in the developing embryo and placenta. Alternatively, it is possible that the observed correlations are due to individual-specific events that occur upstream of, yet contribute to, the establishment of VM-IAP methylation levels. Some VM-IAPs did not exhibit a correlation between tail and placental methylation, which suggests that the underlying mechanisms are not operating simultaneously for all loci. This is illustrated by VM-IAP$_{Fam78b}$: methylation at this locus ranges from 60 to 90% across individuals in the soma, but is substantially lower in pooled blastocysts and individual ESC lines. Together with the lack of correlation between tail and placental methylation, it is likely that VM-IAP$_{Fam78b}$ somatic methylation patterns are deposited later in development compared to other VM-IAPs.

The female germline was only very briefly mentioned in this chapter. The limited number of oocytes that can be harvested from a single B6 female requires the pooling of samples from multiple females. This makes it unfeasible to assess inter-individual differences in methylation or carry out an experiment that matches the one conducted on the male germline. The limited information that can be gained from the pyrosequencing of pooled oocytes can more easily be extracted from publicly available oocyte methylomes. A study using this method was performed by Dr. Hiroyuki Sasaki and Dr. Kenjiro Shirane (Kyushu University, Fukuoka, Japan) and the resulting data were presented in Dr. Anastasiya Kazachenka's doctoral thesis (Kazachenka 2018). The analysis showed that VM-IAPs as a group are neither fully methylated nor completely

unmethylated in the female germline. Instead, VM-IAPs exhibit intermediate methylation of varying degrees and appear to be less resistant to reprogramming when compared to non-variable IAPs (as is the case in the male germline and during preimplantation development). While there is no relationship between *somatic* VM-IAP methylation levels in the mother and those observed in the offspring (with the exception for VM-IAP$_{Gm13849}$, see Chapter 2), it remains possible that the methylation state of the oocyte influences the establishment of methylation levels in the subsequent generation. Of note, VM-IAP$_{Fam78b}$ is lowly methylated in both oocytes and blastocysts, contrasting with the high methylation levels observed in sperm and postnatal somatic tissues.

Additional work is required to more fully understand VM-IAP methylation dynamics during early development. Ongoing efforts by Amir Hay in our group are aimed at uncovering the establishment of variable methylation states *in vitro.* These include clonal expansion experiments in ESCs, cellular differentiation assays, and the use of gastruloids as a model to study early embryogenesis. So far, his results indicate that VM-IAP methylation is not mitotically maintained in ESCs, suggesting these pluripotent cells may be caught in a stochastic stage with the potential to become mitotically heritable upon cellular differentiation. These experiments will further develop our understanding of the developmental dynamics of VM-IAP methylation and will establish robust *in vitro* systems to elucidate the molecular mechanisms governing VM-IAP metastability.

Chapter 4

Environmental modulation of VM-IAPs

4.1 Introduction and objectives

While the *A^vy* mouse model is most widely known as an example of epigenetic inheritance, it has been studied more extensively in the field of environmental epigenetics. As discussed in the introductory chapter, maternal exposure to a range of environmental conditions causes a shift in offspring coat colour distributions, often with a corresponding shift in methylation levels at the *A^vy* IAP (though this has not always been tested). Dietary supplementation of dams with methyl donors and cofactors (e.g., folate, vitamin B_{12}, choline chloride, and anhydrous betaine) leads to a shift in offspring coat colour towards pseudoagouti (Cooney *et al.* 2002; Waterland & Jirtle 2003; Wolff *et al.* 1998). Other pseudoagouti-shifting environmental perturbations include dietary genistein, ethanol, ionizing radiation, and dibutyl phthalates (Bernal *et al.* 2013; Dolinoy *et al.* 2006; Kaminen-Ahola *et al.* 2010; Neier *et al.* 2019). Conversely, perinatal exposures to lead and bisphenol A (BPA) were reported to shift offspring coat colour towards yellow (Dolinoy *et al.* 2007; Faulk *et al.* 2013b), although a later more extensive study was unable to replicate this effect of BPA exposure (Rosenfeld *et al.* 2013). From this, it has been suggested that the *A^vy*

mouse can be used as an epigenetic biosensor of environmental alterations (Dolinoy 2008; Jirtle 2014).

Two other metastable epialleles, *Axin*^{Fu} and *Cabp*^{IAP}, have been analysed in the contexts of methyl donor supplementation and BPA exposure, respectively, and appear to be similarly susceptible (Dolinoy *et al.* 2007; Waterland *et al.* 2006). However, the full extent to which metastable epialleles are environmentally labile with biosensor capabilities is unknown. The assessment of multiple VM-IAPs within and between individuals provides a more predictive and robust test of this question, with the further potential to determine whether VM-IAPs might act as mediators between environmental stimuli and phenotypic outcome. The data presented in this chapter are the product of a multi-collaborative study driven by the following objectives:

1. To determine whether VM-IAP methylation levels are susceptible to:
    a. Abnormal folate metabolism,
    b. Maternal exposure to BPA, or a
    c. Maternal obesogenic diet;

2. If so, to assess whether environmentally-induced VM-IAP methylation changes are associated with altered gene expression;

3. To examine the effect of age on VM-IAP methylation levels.

4.2 Results

4.2.1 Folate metabolism

4.2.1.1 Mtrr^{gt} mouse model

One-carbon metabolism, which encompasses the folate and methionine cycles, is important for epigenetic regulation as it is required for the transfer of methyl groups destined for substrates such as DNA, RNA, and histones (**Figure 4.1**; Friso *et al.* 2017). Normal metabolic progression involves the enzyme methionine synthase (MTR) to exclusively transfer a methyl-group from 5-methyltetrahydrofolate to homocysteine to form methionine and tetrahydrofolate (Shane & Stokstad 1985). In mammals, MTR activity is maintained by methionine synthase reductase (MTRR) through the reductive methylation of its vitamin B12 cofactor (Yamada *et al.* 2006). A hypomorphic mutation in *Mtrr* (*Mtrr^{gt}*) in mice provides a powerful genetic model to study abnormal folate metabolism (Czeizel *et al.* 2013; Elmore *et al.* 2007; Padmanabhan *et al.* 2013). This is because the *Mtrr^{gt}* mutation causes phenotypes commonly associated with dietary folate deficiency in humans (Ducker & Rabinowitz 2017), including increased plasma total homocysteine concentrations (Elmore *et al.* 2007; Padmanabhan *et al.* 2013), macrocytic anemia (Padmanabhan *et al.* 2018), congenital abnormalities (e.g., neural tube, heart, and placenta defects), and locus-specific epigenetic instability (Padmanabhan *et al.* 2013). The work on the *Mtrr^{gt}* mouse line presented below was done in collaboration with Dr. Erica Watson (University of Cambridge). Having developed the *Mtrr^{gt}* mouse line herself, her lab now uses it as a model to study transgenerational epigenetic inheritance. Most of the tissues used were dissected by her former PhD student Gina Blake, as indicated in the figure legends.

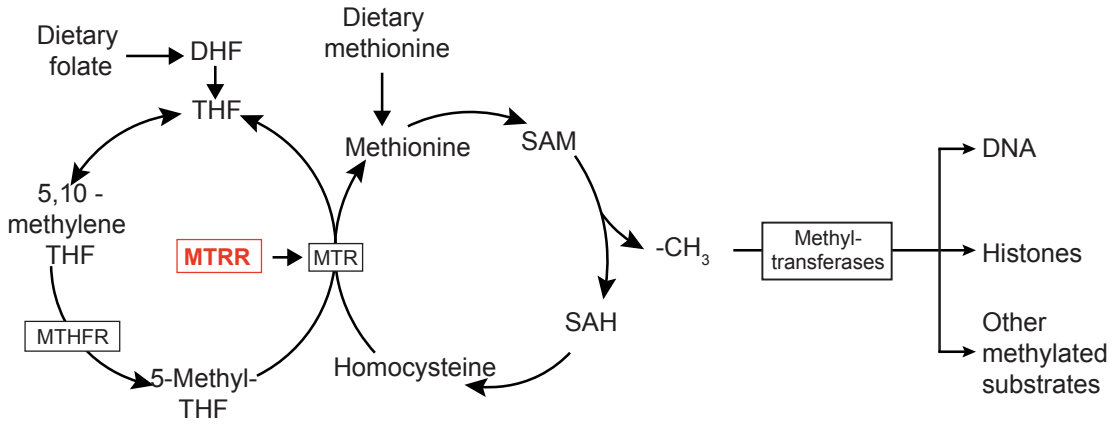


Figure 4.1: Simplified diagram of one-carbon metabolism. The proper functioning of the inter-connected folate and methionine cycles is required for the methylation of biological substrates. Abbreviations: DHF, dihydrofolate; THF, tetrahydrofolate; MTHFR, 5,10-Methylenetetrahydrofolate Reductase; MTR, Methionine Synthase; MTRR, Methionine Synthase Reductase; SAM, *S*-Adenosylmethionine; SAH, *S*-Adenosylhomocysteine. Metabolic enzymes are shown in boxes. The *Mtrr*^{gt} mouse line was generated via a hypomorphic mutation of the *Mtrr* gene, whose product is shown in red in the schematic. This figure was modified from Figure 1A in Padmanabhan *et al.* 2013.

4.2.1.2 VM-IAP methylation states are susceptible to abnormal folate metabolism

To investigate the effect of abnormal folate metabolism on VM-IAP methylation, VM-IAP methylation levels were compared between control and homozygous *Mtrr*$^{gt/gt}$ adult male livers using bisulphite pyrosequencing. Significant differences in methylation levels were observed for eight out of eleven VM-IAP assessed (**Figure 4.2A** and **B**). While most VM-IAPs showed a decrease in methylation in *Mtrr*$^{gt/gt}$ samples compared to controls, two loci (VM-IAP$_{Eps8l1}$ and VM-IAP$_{Rnf157}$) showed an increase in methylation (**Figure 4.2B**). This is consistent with previous reports showing locus-specific CpG hypo- and hypermethylation in *Mtrr*$^{gt/gt}$ tissue (Padmanabhan *et al.* 2013). The magnitude of methylation change was also locus-specific, with VM-IAP$_{Gm13710}$[1], VM-IAP$_{Mbnl1}$, VM-IAP$_{Bmf}$, and VM-IAP$_{Slc15a2}$ showing the largest differences in methylation state compared to controls. With the exception of VM-IAP$_{Eps8l1}$, inter-individual variation in methylation was conserved, and often enhanced, in *Mtrr*$^{gt/gt}$ livers.

[1] VM-IAP$_{Gm13710}$ is referred to as VM-IAP$_{Tfpi}$ in other chapters because VM-IAPs were originally named based on their nearest protein-coding gene (in this case *Tfpi)*. However, the lncRNA *Gm13710* is actually closer. This chapter assesses *Gm13710* expression levels so the name VM-IAP$_{Gm13710}$ is used to avoid confusion.

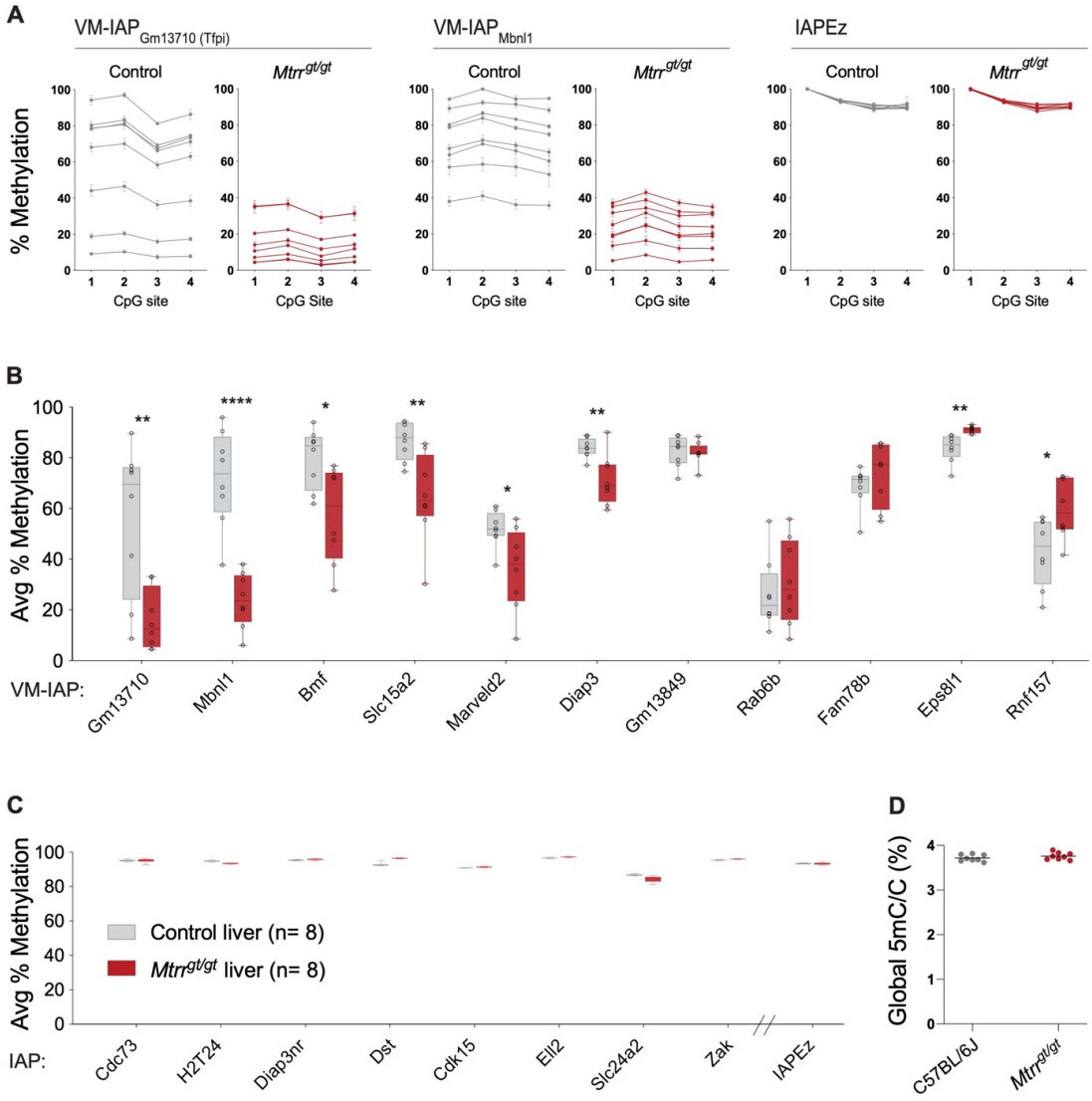


Figure 4.2: Somatic VM-IAP methylation levels are susceptible to abnormal folate metabolism. A. Methylation levels at VM-IAP$_{Gm13710}$, VM-IAP$_{Mbnl1}$, and IAPEz elements in C57BL/6J controls (n = 8, grey lines) and *Mtrr*$^{gt/gt}$ (n = 8, red lines) liver samples. Percent methylation was quantified for the four most distal CpGs of the 5′ LTRs via bisulphite pyrosequencing. **B.** Box plots comparing the average % methylation at eleven VM-IAPs between control (n = 8, grey) and *Mtrr*$^{gt/gt}$ (n = 8, red) liver samples. Reported values represent the average of the four or five most distal CpGs of the VM-IAP 5′ LTRs (Welch's t-tests; * $p < 0.05$; ** $p < 0.005$; **** $p < 0.00005$). **C.** Box plots comparing the average % methylation at non-variable IAPs between control (n = 8, grey) and *Mtrr*$^{gt/gt}$ (n = 8, red) liver samples. Reported values represent the average of the four or five most distal CpGs of the 5′ LTRs. The primers for the bulk IAPEz assay target the LTR of more than 1000 highly conserved elements. **D.** Global DNA methylation levels are equivalent between control (n = 8, grey circles) and *Mtrr*$^{gt/gt}$ (n = 8, red circles) liver samples (unpaired Student's t-test). Global 5-methyl-cytosine (5mC) content was determined by liquid chromatography-tandem mass spectrometry and expressed as a percentage relative to total cytosine (C) in the genome. Box-plot elements: centre line, median; box limits, 25th and 75th percentiles; whiskers, maximum and minimum; all data points shown. The liver tissues used in this analysis were dissected by Gina Blake.

As is the case for all metastable epialleles described to date, VM-IAP methylation levels are variable between individuals but consistent across tissues within an individual (Kazachenka *et al.* 2018). This suggests that the epigenetic status of VM-IAPs is established early in development before tissue differentiation occurs. However, previous work on *Mtrr*$^{gt/gt}$ mice revealed DNA methylation was altered in a tissue-specific manner (Padmanabhan *et al.* 2013), so VM-IAP methylation patterns were explored in *Mtrr*$^{gt/gt}$ brain samples. Results showed that VM-IAP methylation levels at each VM-IAP in *Mtrr*$^{gt/gt}$ brain were comparable to those in *Mtrr*$^{gt/gt}$ liver, indicating that abnormal folate metabolism likely influences the epigenetic state of VM-IAPs during pre-implantation development and prior to differentiation of the germ layers (**Figure 4.3**).

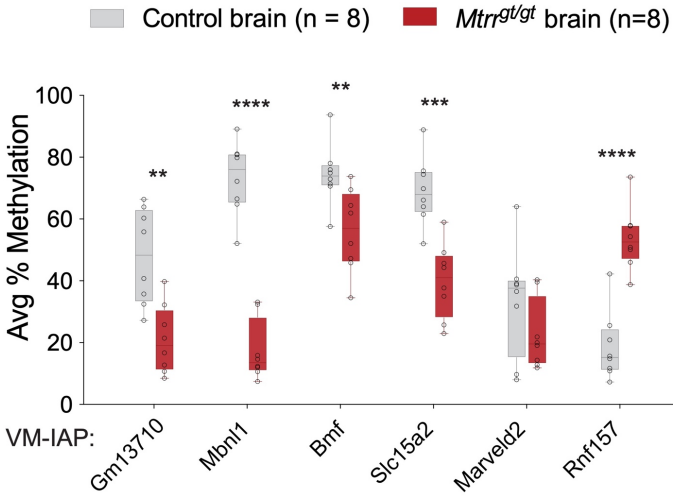


Figure 4.3: VM-IAP methylation levels are altered in *Mtrr*$^{gt/gt}$ brain. Average methylation levels of the four most distal CpGs at VM-IAP 5'LTRs were compared between C57BL/6J control (n = 8, grey box plots) and *Mtrr*$^{gt/gt}$ (n = 8, red box plots) brains (Welch's t-tests; ** $p < 0.005$; *** $p < 0.0005$; **** $p < 0.00005$). Box-plot elements: centre line, median; box limits, 25th and 75th percentiles; whiskers, maximum and minimum; all data points shown. The brain tissues used in this analysis were dissected by Gina Blake.

IAPs that are normally hypermethylated and non-variable in B6 mice remained that way in *Mtrr*$^{gt/gt}$ livers, suggesting that the methylation changes in *Mtrr*$^{gt/gt}$ tissue are specific to VM-IAPs rather than a feature associated with IAP elements in general (**Figure 4.2C**). To ensure these IAP elements represented the epigenetic state of other IAPs in the genome, more than 1,000 highly conserved IAP elements were probed at once using primers targeting the LTR of the IAPEz subclass consensus sequence. This bulk assay confirmed IAP hypermethylation in both control and *Mtrr*$^{gt/gt}$ liver samples (**Figure 4.2C**).

To assess whether altered VM-IAP methylation caused by abnormal folate metabolism reflected a genome-wide hypomethylation phenomenon, global CpG methylation was compared between control and *Mtrr*$^{gt/gt}$ liver DNA using mass spectrometry. No significant differences in global 5-methylcytosine levels were observed between control and mutant livers (**Figure 4.2D**). Together, these findings indicate that the *Mtrr*^{gt} hypomorphic mutation influences DNA methylation in a locus-specific manner and suggest that a stochastically methylated state may render a locus particularly susceptible to abnormal folate metabolism.

4.2.1.3 VM-IAP methylation in the male germline is unaffected by the Mtrr^{gt} *mutation*

Data in Chapter 3 revealed that VM-IAPs are similar to other repeat elements in the male germline, becoming fully methylated during spermatogenesis (**Figure 3.1**; Kobayashi *et al.* 2012). VM-IAP methylation levels were assessed in mature sperm collected from control and *Mtrr*$^{gt/gt}$ adult males to determine whether the effect of the *Mtrr*^{gt} mutation on VM-IAP methylation extends to the male germline. In contrast to somatic tissues, VM-IAP hypermethylation in B6 control sperm was maintained in *Mtrr*$^{gt/gt}$ sperm (**Figure 4.4**). This emphasises that the regulatory mechanisms underlying repeat element methylation in the germline are distinct from the soma. Regulation of repeats is more robust in the male germline, perhaps because the greater evolutionary cost of activating potentially mutagenic mobile elements.

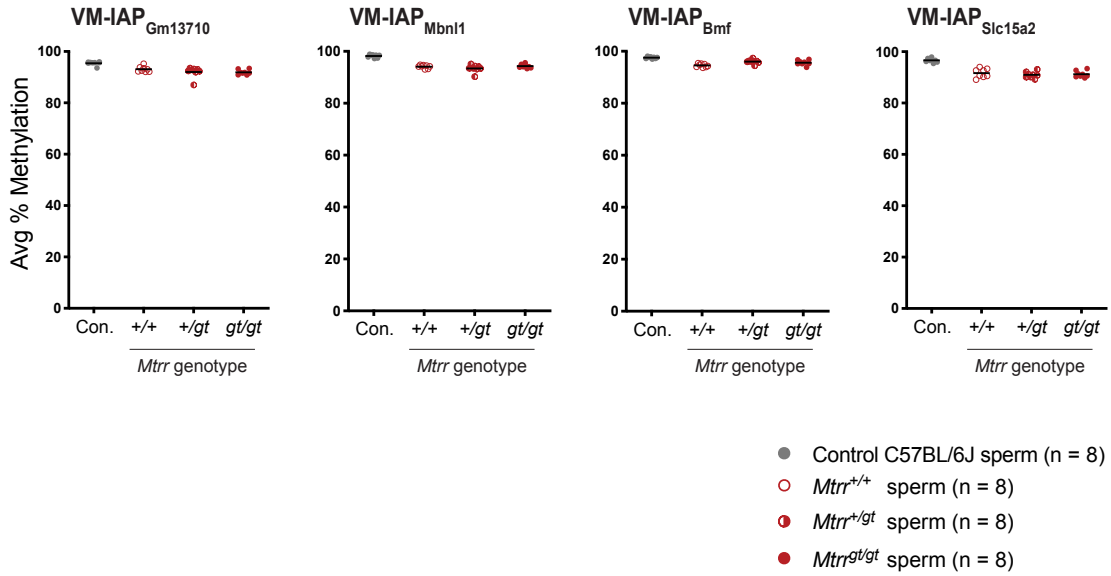


Figure 4.4: VM-IAP methylation levels are unaffected by the *Mtrr*^{gt} allele in mature sperm. Average methylation levels of the four most distal CpGs at VM-IAP 5'LTRs are shown for sperm collected from the cauda epididymides and vas deferens of C57BL/6J control (n = 8, grey circles), *Mtrr*$^{+/+}$ (n = 8, hollow red circles), *Mtrr*$^{gt/+}$ (n = 8, half-filled red circles), and *Mtrr*$^{gt/gt}$ (n = 8, red circles) adult fertile males. The sperm samples used in this analysis were isolated by Gina Blake.

4.2.1.4 *Embryonic VM-IAP methylation states are independent of phenotypes associated with the Mtrr*^{gt} *model*

A wide spectrum of developmental phenotypes appears in *Mtrr*$^{gt/gt}$ conceptuses at midgestation. At embryonic stage E10.5, embryos range from phenotypically normal to severely affected (i.e., display at least one congenital malformation) (Padmanabhan *et al.* 2013). In the first instance, VM-IAP methylation levels were compared between B6 control E10.5 embryos and *Mtrr*$^{gt/gt}$ E10.5 embryos. Both phenotypically normal and severely affected *Mtrr*$^{gt/gt}$ embryos showed significant changes in methylation at each VM-IAP assessed (**Figure 4.5**), which were consistent with the data from adult tissues with regards to degree and directionality of the methylation change (**Figure 4.2**; **Figure 4.5**). This shows that the *Mtrr*$^{gt/gt}$ mutation acts on VM-IAP methylation prior to E10.5 and further supports the hypothesis that abnormal folate metabolism alters the methylation of VM-IAPs at an early stage in development.

This is the earliest developmental time point at which inter-individual methylation variation at VM-IAPs has been reported. Phenotypically normal and severely affected E10.5 *Mtrr*$^{gt/gt}$ embryos did not exhibit significant differences in VM-IAP methylation levels, indicating that the mechanism by which the *Mtrr*^{gt} mutation influences VM-IAP methylation levels is distinct from the process leading to variable phenotypic penetrance in utero (**Figure 4.5**).

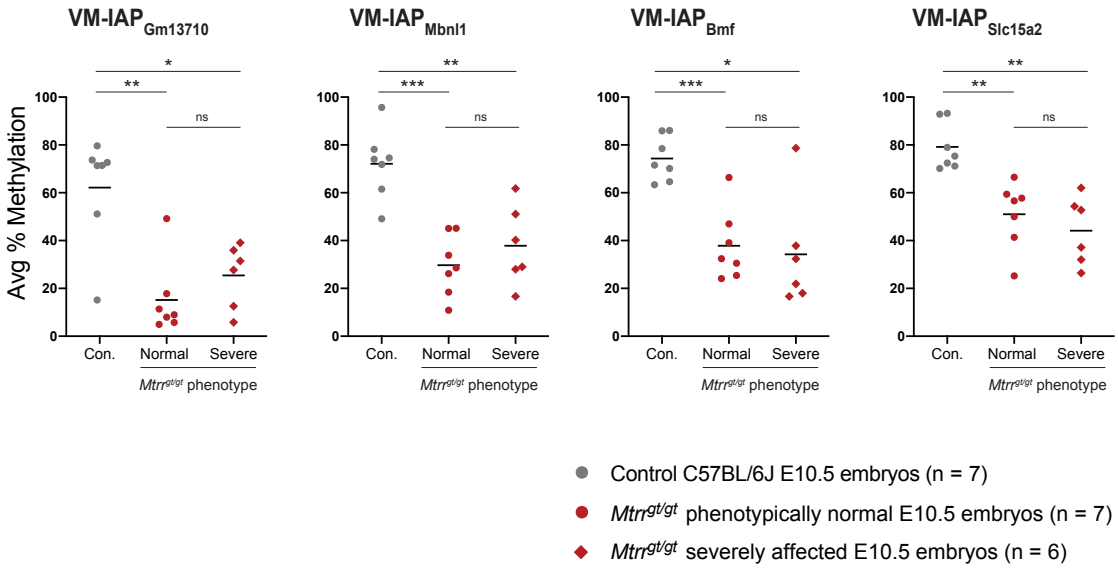


Figure 4.5: VM-IAP methylation states are not associated with phenotypic severity in whole *Mtrr*$^{gt/gt}$ embryos. Average methylation levels of the four most distal CpGs at VM-IAP 5′LTRs were compared for C57BL/6J control embryos (n = 7, grey circles), phenotypically normal *Mtrr*$^{gt/gt}$ embryos (n = 7, red circles), and severely affected *Mtrr*$^{gt/gt}$ embryos (n = 6, red circles) at E10.5 (Welch's ANOVA with Tamhane T2 post hoc tests; * $p < 0.05$; ** $p < 0.005$; *** $p < 0.0005$). The embryos used in this analysis were dissected and phenotyped by Gina Blake and Erica Watson.

Transgenerational epigenetic inheritance is a defining characteristic of the *Mtrr*^{gt} mouse line, making it a model for the study of non-genetic transmission of phenotypic traits (Padmanabhan *et al.* 2013). Maternal grandparents that are carriers of the *Mtrr*^{gt} mutation give rise to wildtype grandprogeny (at least up to the F4 generation) with developmental phenotypes at E10.5 that are similar to those observed in *Mtrr*$^{gt/gt}$ conceptuses (Padmanabhan *et al.* 2013). This transgenerational inheritance of phenotype in the absence of the original *Mtrr*

mutation has been linked to epigenetic instability (Padmanabhan *et al.* 2013). Adult livers from *Mtrr*$^{+/+}$ males derived from *Mtrr*$^{+/gt}$ intercrosses were examined to determine whether the presence of the *Mtrr*^{gt} allele in the previous generation has an effect on the methylation state of VM-IAPs. For all but one of the VM-IAPs tested, *Mtrr*$^{+/+}$ liver methylation levels were comparable to those in B6 control liver (**Figure 4.6**). The exception was VM-IAP$_{Rnf157}$ where mean methylation was significantly decreased in *Mtrr*$^{+/+}$ livers compared to control livers (**Figure 4.6**). This result contrasts with the significant increase in VM-IAP$_{Rnf157}$ methylation levels observed in *Mtrr*$^{gt/gt}$ livers when compared to control livers (**Figure 4.2**). Further experiments are required to resolve whether this finding is biological relevant. VM-IAP methylation levels were also quantified in heterozygous *Mtrr*$^{+/gt}$ livers and no significant changes in methylation between B6 control and *Mtrr*$^{+/gt}$ samples were found (**Figure 4.6**). Altered VM-IAP methylation is therefore largely restricted to *Mtrr*$^{gt/gt}$ individuals and is not affected by parental *Mtrr*^{gt} alleles.

4.2.1.5 Expression of VM-IAP-associated genes is influenced by abnormal folate metabolism

A subset of VM-IAPs provide LTR promoters to adjacent genes in a manner reminiscent of the *A*^{vy} locus, whereby the methylation level at the VM-IAP LTR is inversely correlated with expression (Kazachenka *et al.* 2018; Morgan *et al.* 1999). It was hypothesized that disrupted folate metabolism might influence the behaviour of these VM-IAP-associated genes. Replicating previous experiments (Kazachenka *et al.* 2018), *Slc15a2, Eps8l1,* and *Gm13710* expression levels in B6 liver were inversely correlated with VM-IAP$_{Slc15a2}$, VM-IAP$_{Eps8l1,}$ and VM-IAP$_{Gm13710}$ methylation levels, respectively (r = -0.717, -0.813, -0.929; p-values = 0.048, 0.014, 0.0008, respectively; **Figure 4.7**, left-hand graphs). Combining the datasets from both B6 control and *Mtrr*$^{gt/gt}$ liver samples resulted in stronger correlations between methylation and expression (r = -0.8640, -0.8897, -0.8010; p-values = <0.0001, <0.0001, 0.0002, respectively; **Figure 4.7**,

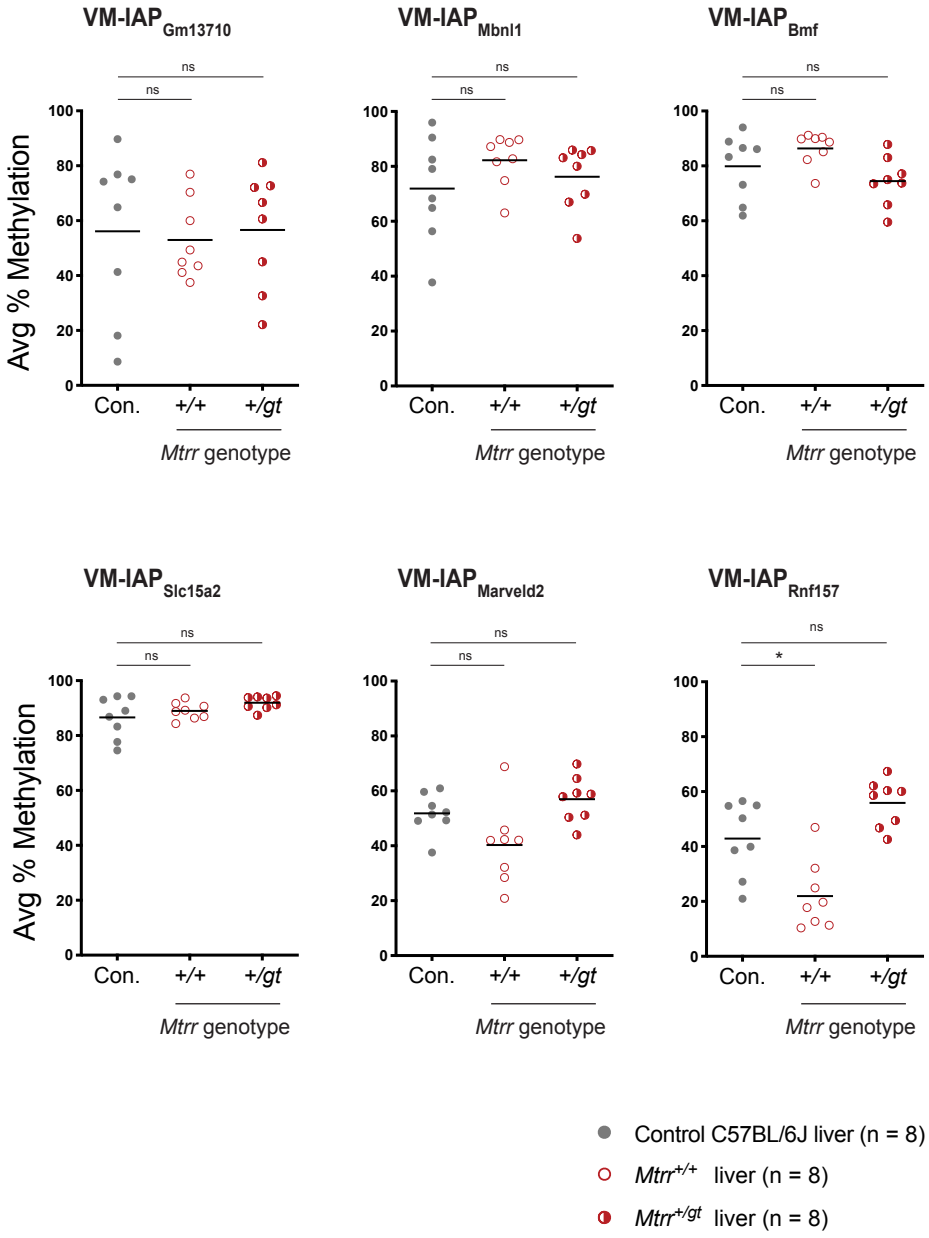


Figure 4.6: Altered VM-IAP methylation levels in the *Mtrr*^{gt} mouse line are restricted to homozygous mutant individuals. Average methylation levels of the four most distal CpGs at VM-IAP 5'LTRs were compared between C57BL/6J control (n = 8, grey circles), *Mtrr*$^{+/+}$ (n = 8, hollow red circles), and *Mtrr*$^{gt/+}$ (n = 8, half-filled red circles) adult liver samples (Welch's ANOVA with Dunnett T3 post hoc tests; * $p < 0.05$; ns, not significant). The liver tissues used in this analysis were dissected by Gina Blake.

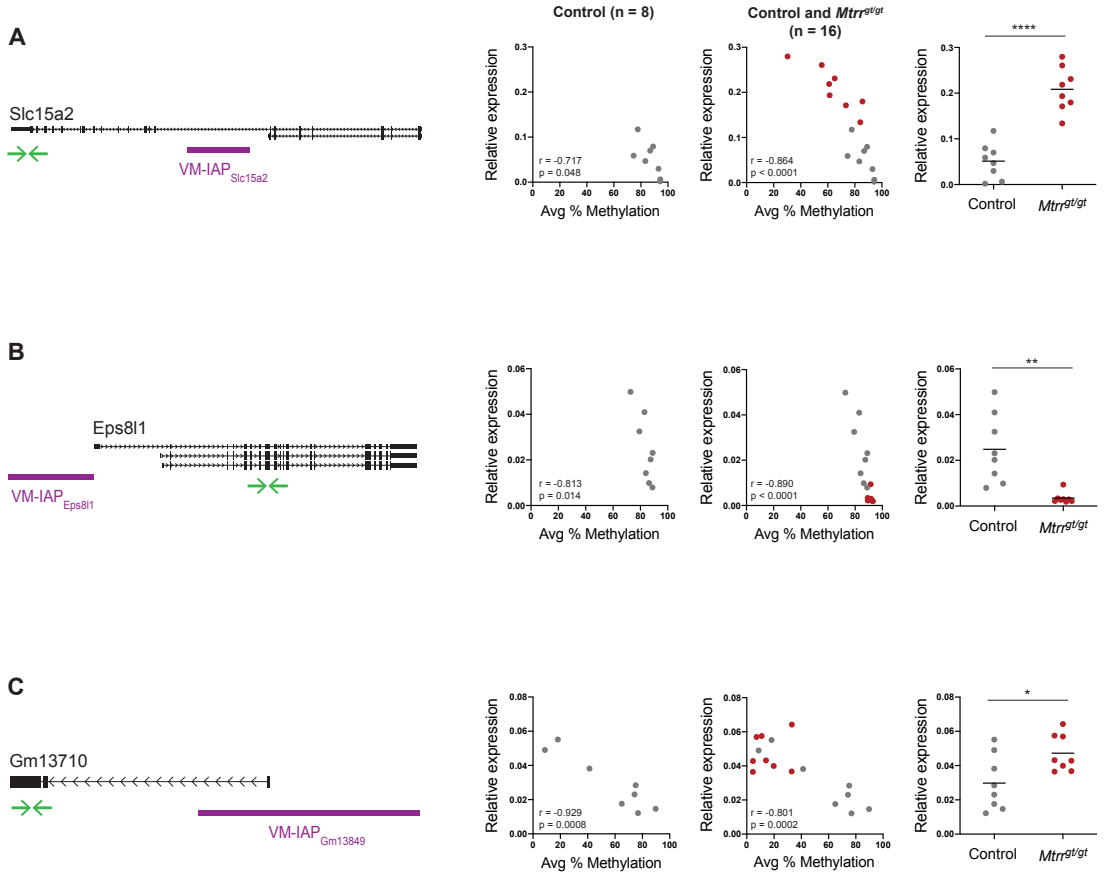


Figure 4.7: VM-IAP-associated gene expression is influenced by abnormal folate metabolism. Right-hand graphs show qRT-PCR expression data of VM-IAP-neighbouring genes *Slc15a2* (**A**), *Eps8l1* (**B**), and *Gm13710* (**C**) in C57BL/6J control (n=8, grey circles) and *Mtrr*$^{gt/gt}$ (n = 8, red circles) liver (Welch's t-tests; * $p < 0.05$; ** $p < 0.005$; **** $p < 0.00005$; ns, not significant). Left-hand graphs assess the correlation between gene expression and VM-IAP methylation in C57BL/6J control livers (n=8, r: Pearson's correlation coefficient; p: two-tailed p-value associated with r). Centre graphs incorporate both control and *Mtrr*$^{gt/gt}$ data and assess the correlation between gene expression and VM-IAP methylation (n = 16, r: Pearson's correlation coefficient; p: two-tailed p-value associated with r). Diagrams of VM-IAPs in relation to their neighbouring gene are depicted on the far left. Gene transcripts extracted from the University of California, Santa Cruz (UCSC) Genome Browser (Haeussler *et al.* 2019) are shown in black and VM-IAPs in purple. Green arrows represent the location of qRT-PCR primers. Diagrams are drawn to scale. The liver tissues used in this analysis were dissected by Gina Blake.

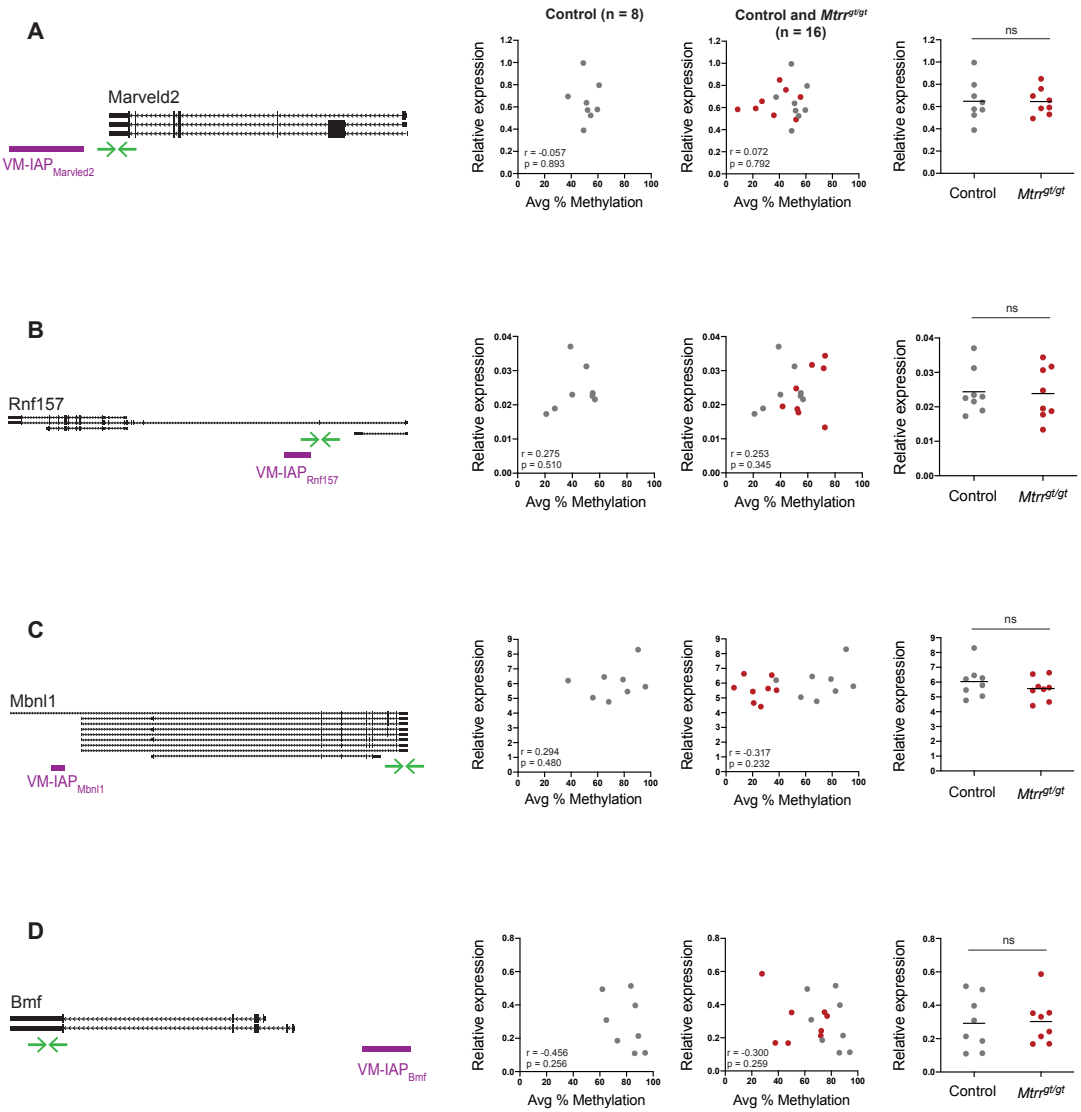


Figure 4.8: Neighbouring genes whose expression is not associated with VM-IAP methylation are not influenced by abnormal folate metabolism. Left-hand graphs show qRT-PCR expression data of VM-IAP-neighbouring genes *Marveld2* (**A**), *Rnf157* (**B**), *Mbnl1*(**C**), and *Bmf* (D) in C57BL/6J control (n=8, grey circles) and *Mtrr*$^{gt/gt}$ (n = 8, red circles) liver (Welch's t-tests; ns, not significant). Centre graphs assess the correlation between gene expression and VM-IAP methylation in C57BL/6J control livers (n=8, r: Pearson's correlation coefficient; p: two-tailed p-value associated with r). Right-hand graphs incorporate both control and *Mtrr*$^{gt/gt}$ data and assess the correlation between gene expression and VM-IAP methylation (n = 16, r: Pearson's correlation coefficient; p: two-tailed p-value associated with r). Diagrams of VM-IAPs in relation to their neighbouring gene are depicted on the far left. Gene transcripts extracted from the UCSC Genome Browser (Haeussler *et al.* 2019) are shown in black and VM-IAPs in purple. Green arrows represent the location of qRT-PCR primers. Diagrams are drawn to scale. The liver tissues used in this analysis were dissected by Gina Blake.

centre graphs), indicating that the relationship between VM-IAP methylation and expression is preserved in altered environments. It is precisely these three genes that showed statistically significant differences in expression between control and *Mtrr*$^{gt/gt}$ liver samples (**Figure 4.7**, right-hand graphs). In contrast, no significant differences in expression were detected for genes whose expression is not associated with VM-IAP methylation (**Figure 4.8**). These results indicate that VM-IAPs can directly mediate changes in gene expression caused by abnormal folate metabolism.

4.2.2 The endocrine disruptor bisphenol A (BPA)

4.2.2.1 BPA exposure model

BPA is a ubiquitous endocrine disrupting chemical mainly found in polycarbonate plastics and epoxy resins. Human exposure to BPA has been linked to metabolic diseases such as type 2 diabetes (Aekplakorn *et al.* 2015; Ahmadkhaniha *et al.* 2014; Sun *et al.* 2014), emphasising the value of developing mouse models to better understand the health implications of this compound. *In utero* exposure to dietary BPA has been reported to shift *A*^{vy} coat colour towards yellow and decrease DNA methylation levels at the 5′ LTR of the *A*^{vy} IAP as well as at the *Cabp*^{IAP} metastable epiallele (Dolinoy *et al.* 2007). A previously characterized mouse model of maternal BPA exposure was used to test whether BPA-induced epigenetic dysregulation is observed at VM-IAPs (Susiarjo *et al.* 2013). The model is associated with male-specific multigenerational (F1, F2) metabolic phenotypes, including increased body fat, glucose intolerance, impaired insulin secretion, and pancreatic islet inflammation (Bansal *et al.* 2017; Susiarjo *et al.* 2015). Some of these effects have been reported in the F3 generation following maternal (F0) exposure to BPA (Bansal *et al.* 2019). In brief, B6 F0 females were fed either a control or BPA-supplemented diet two weeks prior to mating, throughout pregnancy, and during lactation (**Figure 4.9A**). Two different BPA doses representative of human exposure levels were administered: a lower dose of 10 μg/kg/day and an upper dose of 10 mg/kg/day (Susiarjo *et al.* 2013). The latter dose matches that previously fed to *A*^{vy} mice

(Dolinoy *et al.* 2007; Rosenfeld *et al.* 2013). Dr. Rebecca Simmons and Dr. Marisa Bartolomei (University of Pennsylvania) jointly established this model and the study described below was done in collaboration with their labs. Dr. Amita Bansal (Simmons group) and Duy Nguyen (Bartolomei group) performed the DNA extractions on the analysed samples.

4.2.2.2 VM-IAPs are unresponsive to maternal exposure to BPA

VM-IAP methylation levels were quantified in F1 mice exposed to maternal dietary BPA. The exposure period spanned the early stages of embryonic development when VM-IAP methylation levels are likely established. Bisulphite pyrosequencing was performed on DNA from male and female adult livers, assessing the same VM-IAPs analysed in the *Mtrr*^{gt} model. Both males and females were examined due to the sex specificity of phenotypes in this model (Bansal *et al.* 2017; Susiarjo *et al.* 2015). None of the VM-IAPs tested in F1 females showed a significant difference in methylation levels following maternal BPA exposure, regardless of dose (**Figure 4.9B**). Similarly, the majority of VM-IAPs were unaffected in F1 males with the exception of VM-IAP$_{Rnf157}$, which showed a modest but significant increase in methylation in male mice born to dams exposed to the upper BPA dose compared to control mice (**Figure 4.9C**; $p < 0.05$). Of note, VM-IAP$_{Rnf157}$ is the only VM-IAP reported to exhibit a sex effect in Chapter 2, whereby males were more highly methylated than females. Overall, our results indicate that the majority of metastable epialleles tested to date are not affected by maternal exposure to BPA.

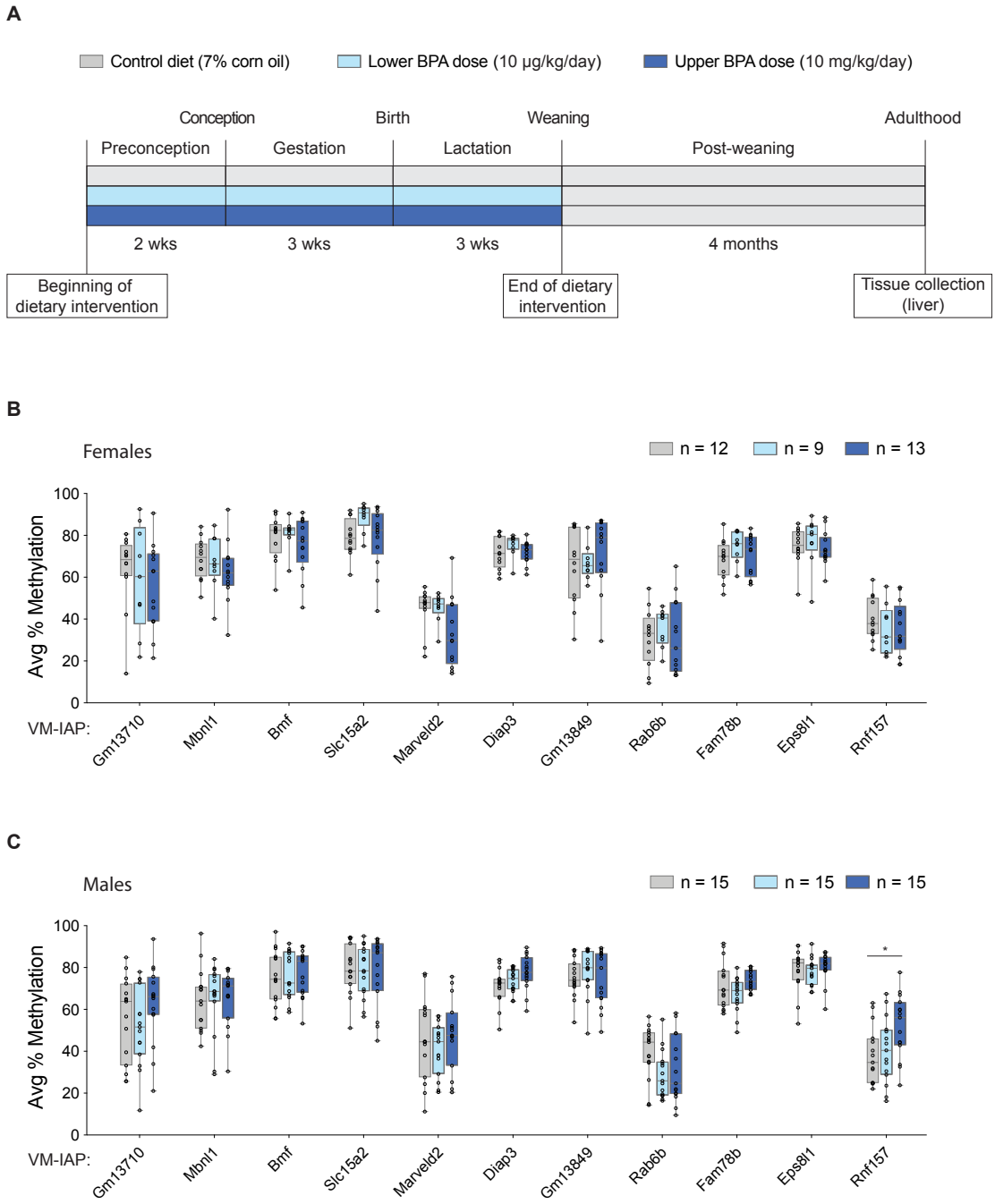


Figure 4.9: VM-IAP methylation is unresponsive to maternal exposure to the endocrine disruptor BPA. A. Scheme of exposure. F0 dams were fed either a control diet (7% corn oil, grey) or one of two BPA-supplemented diets two weeks prior to mating, throughout pregnancy and lactation (lower BPA dose: 10 µg/kg/day, light blue; upper BPA dose: 10 mg/kg/day, dark blue). F1 adult liver tissue was collected from one mouse per litter. **B and C.** The average percent methylation at eleven VM-IAPs in F1 females (A) and F1 males (B) were compared across exposure groups (Welch's ANOVA; Dunnett's T3 post hoc test for VM-IAP$_{Rnf157}$; * $p < 0.05$). Samples sizes are shown in the panel legends. Box-plot elements: centre line, median; box limits, 25th and 75th percentiles; whiskers, maximum and minimum; all data points shown. DNA extractions were performed by Amita Bansal and Duy Nguyen.

4.2.3 Maternal diet-induced obesity

4.2.3.1 Obesogenic diet model

There is substantial evidence that epigenetic mechanisms contribute to the regulation of molecular events governing obesity. Although metastable epialleles have not been investigated specifically in the context of obesity-inducing diets, significant changes to the epigenome have been reported following this type of dietary intervention (Andersen *et al.* 2019; Lavebratt *et al.* 2012; Radford 2018). Therefore, VM-IAP methylation levels were examined in the context of maternal exposure to an obesogenic diet using a well-established mouse model (Samuelsson *et al.* 2008). Specifically, B6 females were fed either a standard chow diet or a calorie-rich obesogenic diet ten weeks before the experimental mating, throughout pregnancy, and during lactation (**Figure 4.10A**; Samuelsson *et al.* 2008). A wide range of cardiovascular and metabolic abnormalities are observed in F1 offspring born to obese dams (Alfaradhi *et al.* 2016; Loche *et al.* 2018; Samuelsson *et al.* 2008). This model was developed by Dr. Susan Ozanne (University of Cambridge) and this portion of the study was carried in collaboration with her lab. Dr. Denise Fernandez-Twinn, a postdoctoral associate in the Ozanne group, dissected the liver tissues used in the experiment described below.

4.2.3.2 Exposure to a maternal obesogenic diet has no effect on VM-IAP methylation

We tested VM-IAP methylation levels (as above) in adult liver samples from F1 males of dams fed the control diet or obesogenic diet and found no significant differences between groups (**Figure 4.10B**). These findings combined with those from the BPA exposure model suggest that metastable epialleles are more robust and resistant to environmental changes in early life than previously suggested (Jirtle 2014).

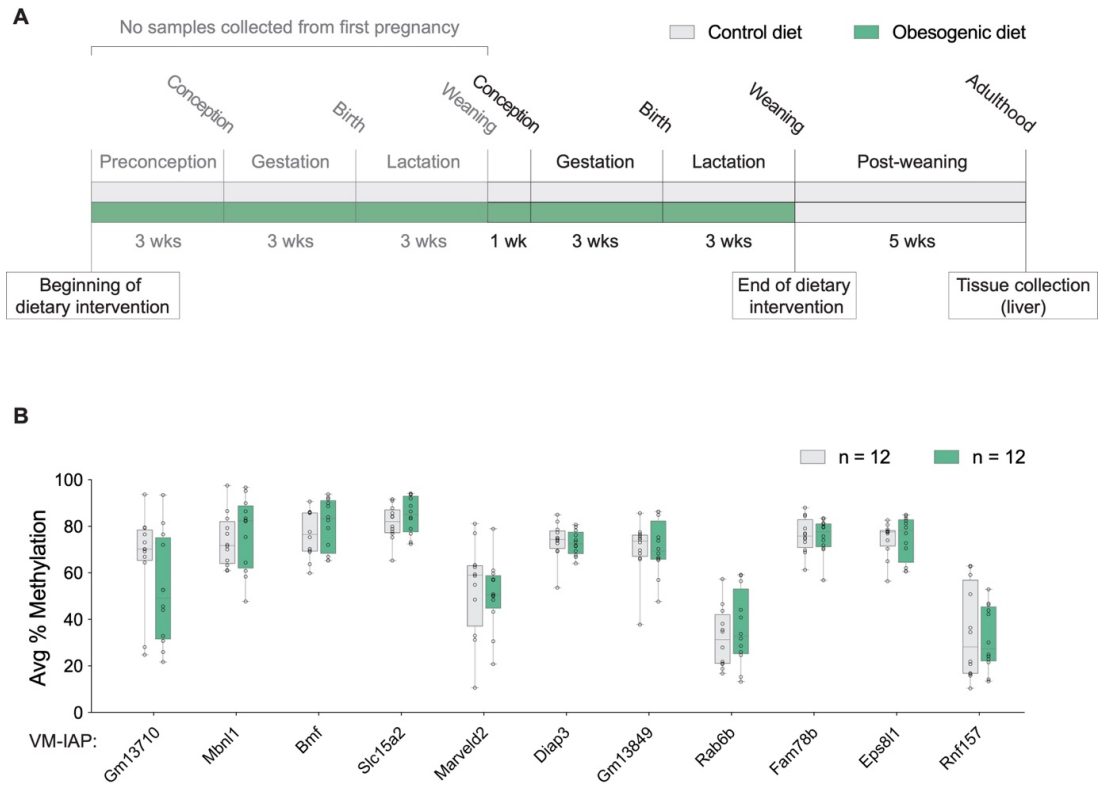


Figure 4.10: Maternal exposure to an obesogenic diet has no effect on VM-IAP methylation levels. A. Scheme of exposure. F0 dams were fed either a control standard chow (grey) or an obesogenic diet (green) ten weeks prior to the experimental mating, throughout pregnancy, and during lactation. The first pregnancy was performed to prove fertility and establish good maternal care. F1 male liver tissue from the second pregnancy was collected from one mouse per litter. **B.** The average percent methylation at eleven VM-IAPs was compared between the two diet groups and no significant differences were found (Welch's t-tests). The values shown on the box plots represent the average of the four or five most distal CpGs of the VM-IAP 5′ LTRs. Samples sizes are shown in the panel legend. Box-plot elements: centre line, median; box limits, 25th and 75th percentiles; whiskers, maximum and minimum; all data points shown. The liver tissues used in this analysis were dissected by Dr. Denise Fernandez-Twinn.

4.2.4 Ageing

4.2.4.1 VM-IAP methylation levels are stable across the murine lifespan.

As shown in Chapter 2, offspring born to a single inbred mouse exhibit a wide range of VM-IAP methylation levels regardless of parental methylation state. This remarkable cross-generational variability in epigenetic state, or metastability, is perhaps the most defining feature of metastable epialleles. Furthermore, the consistency of methylation levels across tissues within a single mouse highlights the stability of VM-IAP methylation states after their establishment. However, the extent to which VM-IAP methylation levels change throughout the murine lifespan has never been analysed. This is of interest in view of the so-called "DNA methylation clocks" developed in the past decade that establish DNA methylation as a useful age predictor in humans (Bocklandt *et al.* 2011; Hannum *et al.* 2013; Horvath 2013) and mice (Meer *et al.* 2018; Petkovich *et al.* 2017; Stubbs *et al.* 2017; Wang *et al.* 2017).

A longitudinal ageing study was developed to track VM-IAP methylation levels *in vivo* in 20 individual B6 mice. Given the inter-individual methylation variation characteristic of VM-IAPs, it was important to assess the same mice over time rather than taking a cohort-based approach with separate age-sorted groups of mice. Ear notch samples were taken at regular 32-week intervals starting 10 days after birth until death. With the exception of one mouse, lifespan ranged from 20 to 31 months, with an average of 26 months. Linear mixed-effects models (LMMs) were used to determine whether age had an effect on VM-IAP methylation level while accounting for the non-independence of littermates and of data points collected from the same mouse across time. The model for each VM-IAP included age as a fixed effect and mouse and litter as random intercept effects. The effect size at each VM-IAP was assessed by measuring marginal R^2 values (R^2m), ranging from 0 to 1 and representing the proportion of methylation variability explained by age.

The analysis revealed that age does not have a profound effect on VM-IAP methylation levels, illustrated by weak effect sizes of 0 to 0.12 (**Figure 4.11**;

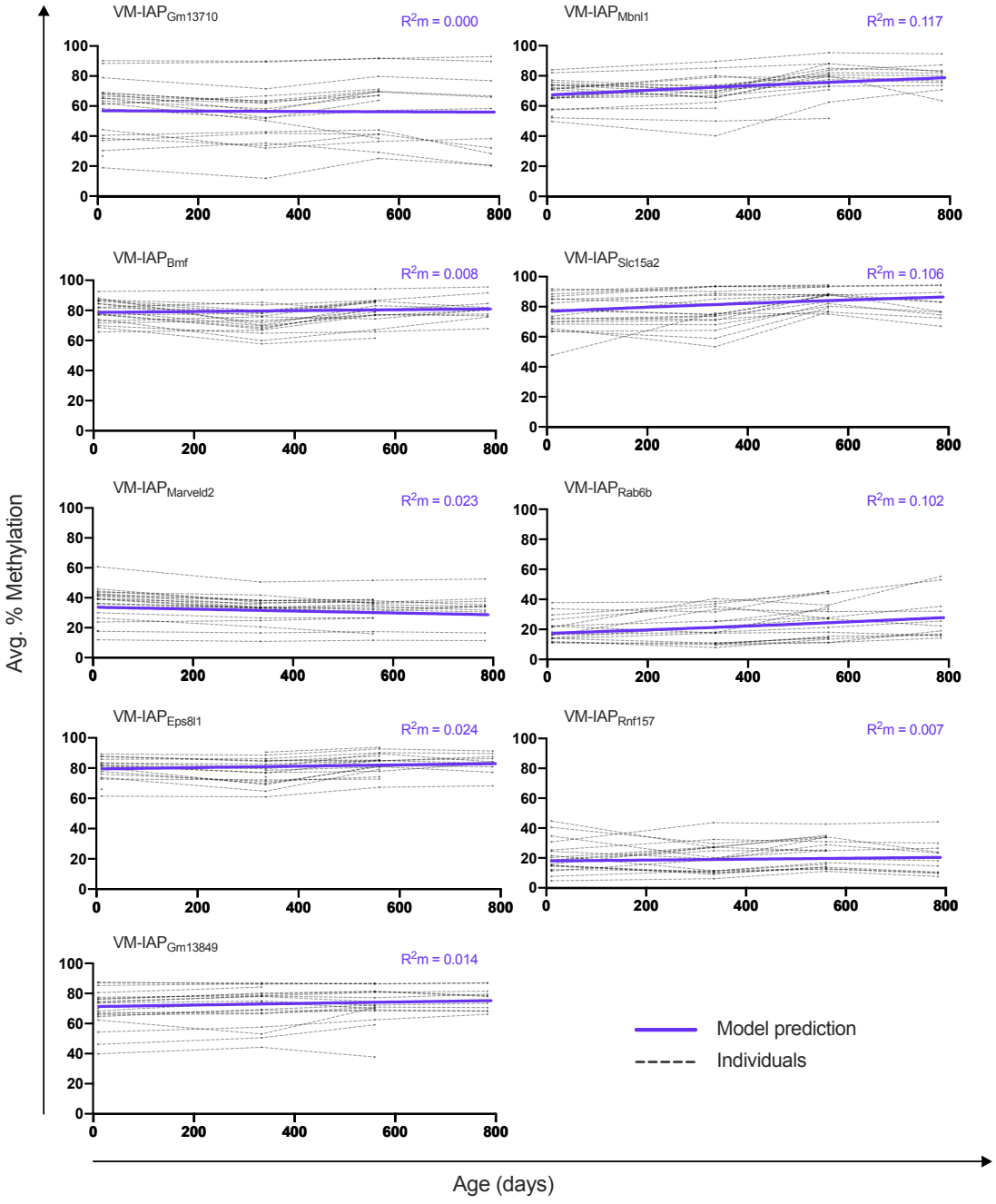


Figure 4.11: VM-IAP methylation levels are stable throughout the murine lifespan. Linear mixed-effects models (LMMs) were built to assess the effect of age on VM-IAP methylation level. Each VM-IAP model was fitted using the lmer() function from the lme4 R package. Age was included as a fixed effect, litter and mouse ID as random intercept effects. Marginal R squared (R^2m) values were calculated to quantify the proportion of variance explained by age. Each dotted line represents VM-IAP methylation of one individual mouse over time. The solid purple line is the model prediction.

Table 4.1: LMM summary statistics from longitudinal ageing study.

VM-IAP	Estimate	Standard error	t value	R2m	p-value*
VM-IAP$_{Gm13710}$					
Intercept	0.299	0.392	0.763		
Age(days)	-4.67E-05	1.15E-04	-0.408	0.000	0.685
VM-IAP$_{Mbnl1}$					
Intercept	0.736	0.125	5.887		
Age(days)	0.001	1.40E-04	5.383	0.117	2.13E-06
VM-IAP$_{Bmf}$					
Intercept	1.305	0.122	10.731		
Age(days)	1.86E-04	1.32E-04	1.408	0.008	0.165
VM-IAP$_{Slc15a2}$					
Intercept	1.224	0.158	7.738		
Age(days)	0.001	1.39E-04	6.108	0.106	1.70E-07
VM-IAP$_{Marveld2}$					
Intercept	-0.668	0.154	-4.346		
Age(days)	-2.95E-04	4.73E-05	-6.242	0.023	1.14E-07
VM-IAP$_{Gm13849}$					
Intercept	0.914	0.133	6.890		
Age(days)	2.57E-04	5.94E-05	4.320	0.014	7.96E-05
VM-IAP$_{Rab6b}$					
Intercept	-1.565	0.139	-11.256		
Age(days)	0.001	0.000	6.671	0.102	3.46E-08
VM-IAP$_{Eps8l1}$					
Intercept	1.364	0.115	11.851		
Age(days)	2.95E-04	1.02E-04	2.883	0.024	0.006
VM-IAP$_{Rnf157}$					
Intercept	-1.495	0.140	-10.689		
Age(days)	1.94E-04	1.34E-04	1.442	0.007	0.156

*Bonferroni-adjusted α-value = 0.0056.

Table 4.1). Despite the small effect sizes, VM-IAP$_{Mbnl1}$, VM-IAP$_{Slc15a2}$, and VM-IAP$_{Rab6b}$ showed significant increases in methylation over time (R^2m = 0.117, 0.106, 0.102, respectively; p-values = 2.13e-6, 1.70e-7, 3.46e-8, respectively; Bonferroni-adjusted α value = 0.0056; **Figure 4.11**; **Table 4.1**), which may indicate that a subset of VM-IAPs have a tendency to gain methylation with age. VM-IAP$_{Marveld2}$ showed a significant decrease in methylation over time (R^2m = 0.023, p-value = 1.14e-7, Bonferroni-adjusted α value = 0.0056; **Figure 4.11**; **Table 4.1**). This behaviour is unlikely to be biologically significant given the small effect size. Importantly, the inter-individual methylation variation of VM-IAPs is considerably greater than the modest absolute change over time, indicating that these loci cannot be used to predict age.

Tracking the same set of mice over time showed that mice that were at the high end of the VM-IAP methylation range at birth stayed in that relative position throughout their life, and vice versa for mice born at the low end of the range (**Figure 4.11**; **Table 4.1**). This indicates that the rank order in methylation of each individual in relation to other individuals remains constant at specific VM-IAPs and that VM-IAP methylation levels are faithfully maintained over the murine lifespan.

4.2.4.2 Cohort-based validation of longitudinal study

The longitudinal nature of the ageing experiment presented above necessitated the use of a tissue type that can be sampled repeatedly in the same individual, hence the use of ear biopsies for the analysis. However, this experimental design does not account for a possible age dependency in more proliferative internal organs. This was explored using a cohort-based approach in collaboration with Dr. Anne Corcoran (Babraham Institute). VM-IAP methylation levels were quantified in liver tissue collected from 6- and 24-month-old B6 male mice housed in the same conditions throughout their life. In line with the results from the longitudinal study, VM-IAP methylation levels were not significantly different between the two age cohorts (**Figure 4.12**). VM-IAP$_{Eps8l1}$ was the only locus to show a significant difference, with 24-month old livers exhibiting greater methylation levels than 6-month old livers (**Figure 4.12**; p = 0.0246). This finding requires independent replication considering no

ageing effect was detected at this region in ear samples and no other VM-IAP displayed this behaviour.

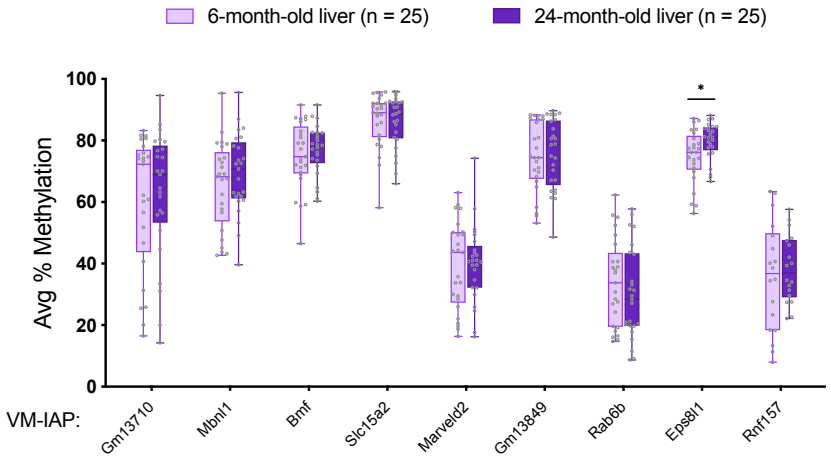


Figure 4.12: 6- and 24-month old liver samples exhibit comparable VM-IAP methylation levels. Liver tissue was dissected from 6- and 24-month old B6 male mice. The average percent methylation at nine VM-IAPs was compared between the two age groups and no significant differences were found for eight of them (Welch's t-tests). VM-IAP$_{Eps8l1}$ exhibited significantly greater methylation levels in 24-month old liver samples ($p = 0.0246$, Welch's t-test). Samples sizes are shown in the panel legend. Box-plot elements: centre line, median; box limits, 25th and 75th percentiles; whiskers, maximum and minimum; all data points shown as grey circles.

4.2.4.3 VM-IAP methylation levels are altered in tumour samples

Seven of the 20 B6 mice used in the longitudinal ageing study were culled due to the development of one or more spontaneous tumours, which is relatively common in aged inbred mice. Given that DNA methylation patterns are often altered in cancerous samples (Klutstein *et al.* 2016), these tumours provided material to conduct an opportunistic and very preliminary analysis to determine whether tumourigenesis has an effect on VM-IAP methylation levels. Tumour samples were only collected from unmistakeable hard white masses. Whenever possible, matched healthy-looking samples of the same tissue type were also harvested. In one instance (Indv. 6 in **Figure 4.13**), the whole liver was diseased

so a healthy kidney sample was harvested instead. One mouse harboured growths in the kidney, lung, colon, and liver (Indv. 1 in **Figure 4.13**); the other six individuals only had one obvious tumour.

The methylation state of VM-IAPs appeared dysregulated in every matched sample pair examined, but not at every locus (**Figure 4.13**). The affected VM-IAPs varied across individuals and there was no consistency in directionality or effect size (**Figure 4.13**). Whether this is due to differences in tumour types has yet to be determined. It is possible that the altered VM-IAP methylation levels are a reflection of 'stemness', whereby the pluripotent nature of tumour cells reverts VM-IAPs to a more labile state. The data generated from the four matched sample pairs collected from Indv. 1 were particularly informative. Methylation levels were constant across the four healthy somatic tissues at each VM-IAP, reinforcing previous findings on metastable epialleles (**Figure 4.13**). In contrast, the tumour samples showed altered methylation levels at almost every tested VM-IAP and exhibited much greater cross-tissue variability in methylation levels (**Figure 4.13**). Of note, the order of methylation levels in tumour tissues was consistent across VM-IAPs: the liver tumour consistently deviated the farthest from the control samples and the kidney tumour consistently deviated the least. This result is highly unusual, as VM-IAP methylation does not co-vary in healthy individuals (Chapter 2). It is tempting to theorise that the tissue-specific order of tumour methylation levels is a reflection of cancer metastasis. Perhaps the cancerous growth originated in the liver and metastasised to other organs, first to the colon, then the lung, and finally the kidney. If this were the case, VM-IAPs could prove useful in tracing the spread of cancer across the body.

No statistical tests could be carried out for this preliminary analysis because the nature of the dissected tumours was unknown and the tissue type varied across individuals. Nonetheless, the observations provide evidence that VM-IAP methylation levels are influenced by tumourigenesis. Additional studies using well-established mouse models of various cancer types will be more informative in determining the extent to which VM-IAP methylation levels are altered in this context. If this finding is confirmed, follow-up studies aimed at understanding the functional consequences of such dysregulation and the mechanisms underlying it could have implications for human health.

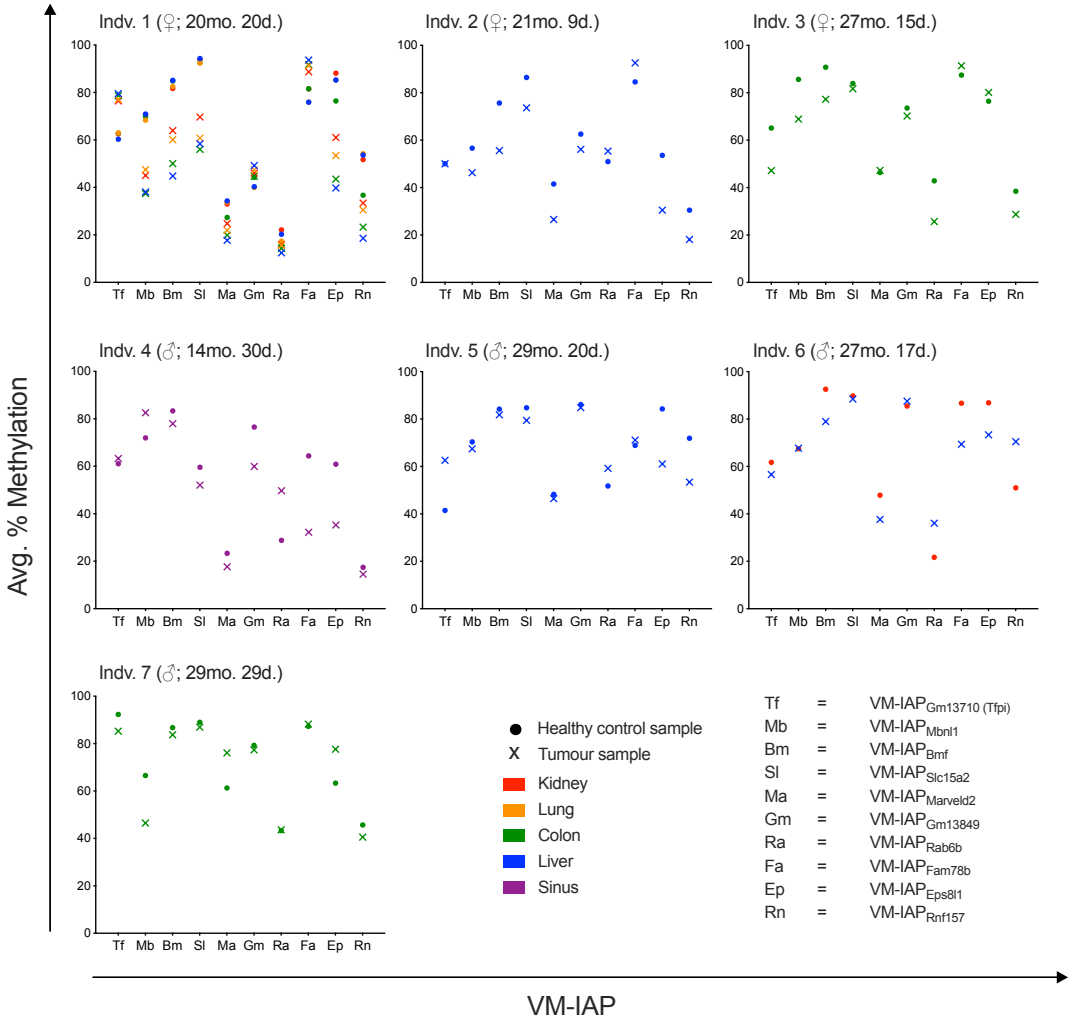


Figure 4.13: Preliminary evidence suggests tumourigenesis alters VM-IAP methylation levels. Tumours (crosses) and matched healthy samples of the same tissue type (circles) were harvested from seven aged mice. Indv. 1 harboured tumours in four tissues; the six others harboured a single tumour. The average percent methylation at 10 VM-IAPs is shown. Tissue types are colour-coded and shown in the legend. Each graph shows data for all VM-IAPs from samples collected from one individual. The sex and age at death of the individual is indicated above each graphs (mo. = months; d. = days).

4.3 Discussion

Over the past two decades, the A^{vy} mouse model has been considered a compelling paradigm both for transgenerational epigenetic inheritance and as an epigenetic sensor of environmental insults. The experiments described in this chapter examined VM-IAP methylation states in three different mouse models of environmental compromise, all of which are associated with multi-generational effects and have implications for human health and disease.

Results showed that somatic VM-IAP methylation levels were particularly responsive to abnormal folate metabolism caused by *Mtrr*$^{gt/gt}$ homozygosity when compared to the rest of the genome and to non-variable IAPs. This increased sensitivity to perturbed folate metabolism is specific to VM-IAPs since other non-variable members of this class are spared. It is possible that metastable epialleles are not recognised as 'priority loci' by the methylation machinery in situations where methyl-group availability is limited. Alternatively, their hypomethylation might reflect potential adaptive benefits of stochastic epigenetic variability, with labile methylation conferring fitness by more readily responding to fluctuating environments.

The size and directionality of the methylation effects detected in *Mtrr*$^{gt/gt}$ mice were locus-specific, suggesting that the mechanisms governing the establishment of inter-individual methylation variability at VM-IAPs are, at least in part, also locus-specific. The increase in methylation observed in *Mtrr*$^{gt/gt}$ mice at VM-IAP$_{Eps8l1}$ and VM-IAP$_{Rnf157}$ indicates that the epigenetic consequences of *Mtrr* deficiency are complex and cannot be explained by a simple depletion in available methyl groups, as this would only result in decreases in methylation. This is consistent with the lack of change in global DNA methylation levels in *Mtrr*$^{gt/gt}$ livers and in line with previously published work showing locus-specific hypo- and hypermethylation in *Mtrr*$^{gt/gt}$ placentas (Padmanabhan *et al.* 2013). All VM-IAPs remained hypermethylated in *Mtrr*$^{gt/gt}$ sperm, reflecting the importance of epigenetic repression of such elements in the male germline.

VM-IAP-associated genes displayed altered expression levels in *Mtrr*$^{gt/gt}$ mice. This has functional implications, as it suggests that VM-IAPs are capable of mediating changes in gene expression in response to abnormal folate

metabolism. For those VM-IAPs that are not adjacent to genes, it remains to be determined whether altered VM-IAP methylation levels in *Mtrr*$^{gt/gt}$ mice have an effect on gene expression via long-range *cis*-interactions. This is an intriguing possibility given that VM-IAP bordering regions are enriched in CCCTC binding factor (CTCF), an important regulator of 3D structure and long-range interactions in the genome (Kazachenka *et al.* 2018; Phillips & Corces 2009).

Unlike the other two investigated mouse models, the *Mtrr*$^{gt/gt}$ mouse model uses a genetic approach to incur metabolic effects similar to folate deficiency, avoiding the complex pathways linking diet and metabolism. Indeed, *Mtrr*$^{gt/gt}$ mice recapitulate the phenotypes observed in humans with folate deficiency, while diet-induced folate deficient mice do not unless they are challenged with genetic insufficiency (Ducker & Rabinowitz 2017; Padmanabhan *et al.* 2013, 2018). However, further experiments are required to determine whether a dietary model of folate deficiency would result in comparable epigenetic dysregulation of VM-IAPs.

In addition, a more in-depth characterisation of the genetic context of the *Mtrr* mutation in this mouse line is required. The *Mtrr*^{gt} allele was originally generated in the 129P2/OlaHsd (129/Ola) mouse strain. Subsequent backcrossing to the C57BL/6J strain for eight generations yielded the mouse line used for the experiments described in this chapter. However, it is highly probable that the segments of DNA on either side of the *Mtrr*^{gt} mutation are still 129/Ola in origin. Given that the *Mtrr* gene is located next to a cluster of KZFPs (important epigenetic regulators of ERVs) and that Chapter 5 of this thesis demonstrates that VM-IAPs are subject to genetic background effects, it is crucial to determine the size of the 129/Ola interval and investigate the potential role of genes lying within this region in modulating VM-IAP methylation. This opens up the interesting possibility that genetic variation at modifier loci may underlie apparent environmental effects on VM-IAPs.

Despite the similarities in dosage and exposure window between the present study and the one by Dolinoy and colleagues conducted on the *A*^{vy} locus (Dolinoy *et al.* 2007), no BPA-associated effect on VM-IAP methylation levels was found, except a modest increase at VM-IAP$_{Rnf157}$ in male offspring of dams exposed to the upper BPA dose. In this regard our findings are more consistent

with a later study showing that A^{vy} coat colour is unaffected by maternal exposure to BPA (Rosenfeld *et al.* 2013). It is possible that statistically significant differences would have been detected with larger sample sizes, but these methylation changes would have been associated with small effect sizes lacking functional importance. In addition, since effective biosensors necessitate large effect sizes, metastable epialleles are poor epigenetic biosensors of environmental compromise.

The longitudinal ageing study indicated that VM-IAPs are generally insensitive to age, with only three VM-IAPs showing slight increases in methylation over time. This renders VM-IAPs inadequate for age prediction. Importantly, the rank order in methylation of individuals at each locus did not change. This implies that the signature methylation stochasticity at VM-IAPs is linked to early developmental events, after which point the epigenetic state is maintained. It is this early window of preimplantation development that has been targeted in most studies investigating the environmental modulation of A^{vy} and other metastable epialleles. The work presented in this chapter adds multiple new loci to the list of tested regions and demonstrates that the environmental susceptibility of metastable epialleles is region-specific, with a potential increased sensitivity to perturbed folate metabolism than to other exposure types. As more loci are assessed under a range of different conditions, we will gain a broader understanding of the degree to which stochastic epigenetic states are hotspots for environmental vulnerability. These analyses have implications for health and disease, as they will determine whether such regions can act as mediators of phenotypic outcomes.

Chapter 5

Strain-specific modification of VM-IAPs

5.1 Introduction and objectives

Parent-of-origin effects arise when the phenotypic outcome of a gene is dependent on the parent it was inherited from. The best-described example is genomic imprinting, the epigenetically-controlled process leading to the differential expression of the maternal and paternal alleles (Ferguson-Smith 2011). Other common instances of parent-of-origin effects involve maternal effects, including intrauterine environment, oocyte cytoplasmic factors, and mitochondrial inheritance (Guilmatre & Sharp 2012). The metastable epialleles A^{vy} and $Axin^{Fu}$ display parent-of-origin effects in that A^{vy}/a mice are 15% more likely to be yellow if the A^{vy} allele is transmitted maternally and $Axin^{Fu}/+$ mice are 30% more likely to have kinked tails if the $Axin^{Fu}$ allele is inherited paternally (Blewitt & Whitelaw 2013). The mechanisms driving these effects are not well understood.

Intrinsic differences between the maternal and paternal alleles are difficult to detect in inbred mice because their genetic homogeneity prevents parental lineage tracing. Reciprocal hybrid breeding schemes are used in mouse genetics as a powerful tool to study parent-of-origin effects. The crossing of two inbred mouse strains results in F1 hybrids whose genomes contain equal and identical

contributions of each parental strain, with single nucleotide polymorphisms (SNPs) and other genetic differences between the two strains enabling the differentiation of alleles. In the case of classically imprinted loci, the F1 allele inherited from one parental strain is methylated while the F1 allele inherited from the other parental strain is unmethylated. In the reciprocal cross, the effect is inversed.

Reciprocal hybrid crosses can also reveal genetic background effects, which emerge when molecular or phenotypic outcomes are modulated by genetic background (i.e. mouse strain). These effects manifest when F1 hybrid offspring exhibit a difference in phenotype compared to the pure parental strains regardless of breeding direction. Reciprocal breeding of inbred 129P4/RrRk and C57BL/6J (B6) mice carrying the A^{vy} and $Axin^{Fu}$ loci, respectively, has shown that the heritability of these alleles is influenced by genetic background (Rakyan *et al.* 2003). Of note, parent-of-origin and genetic background effects are not mutually exclusive, as illustrated by instances where molecular or phenotypic outcomes are dependent on the genetic background of one of the parents (more commonly that of the mother).

The experiments described in this thesis so far have been restricted to the B6 genetic background. Characterising parent-of-origin and genetic background effects has the potential to provide mechanistic insight into epigenetic variation through the mapping and subsequent identification of modifying factors. The work presented in this chapter introduces genetic variation to the study of VM-IAPs and in doing so investigates the genetic contributions to VM-IAP metastability. The objectives are two-fold:

1. To determine whether VM-IAP methylation distributions are susceptible to parent-of-origin and genetic background effects;

2. To establish whether these effects can be harnessed to identify genetic modifiers of VM-IAPs.

5.2 Results

5.2.1 Analysis of parent-of-origin and genetic background effects in F1 hybrid mice

5.2.1.1 Study design

A reciprocal hybrid breeding scheme was implemented using B6 and inbred wild-derived CAST/EiJ (CAST) mice (**Figure 5.1**). These strains were selected for both practical and biological reasons. Practically, the VM-IAPs studied in this thesis are present in B6 and absent from CAST (Kazachenka *et al.* 2018), so F1 hybrid offspring inherit a single VM-IAP copy from their B6 parent. This avoids relying on SNPs to determine the parental origin of the VM-IAP allele and simultaneously allows VM-IAP methylation to be assessed in a hemizygous context, which has not been done before. From a biological point of view, B6 and CAST mice represent evolutionarily distant subspecies, so they are more likely to have evolved different VM-IAP modifiers.

The study design and potential experimental outcomes are illustrated in **Figure 5.1** and described in detail in the corresponding figure legend. Maternal and paternal transmission of each of 12 VM-IAPs was achieved by crossing B6 females to CAST males (BC) and CAST females to B6 males (CB), respectively. Standard nomenclature for mouse genetics crosses was adhered to throughout: the abbreviation of the female parent is listed first, followed by the abbreviation of the male parent. At least 30 mice were analysed per genetic background, with more than 100 mice assessed for some VM-IAPs. Genomic DNA was extracted from postnatal day 10 ear samples and DNA methylation levels were quantified by bisulphite pyrosequencing. Methylation concordance across CpGs at each VM-IAP was not disrupted in hybrids, so CpG methylation values were averaged at each locus before further analyses. B6 frequency distributions were previously only described for six VM-IAPs (see **Figure 2.10** in Chapter 2), so these were generated for the remaining six VM-IAPs for the comparative purposes of this chapter. These appear in **Figure 5.2** and in Appendix A (**Figure A.4**, **Figure A.5**, **Figure A.6**, and **Figure A.7**), confirming that VM-IAP

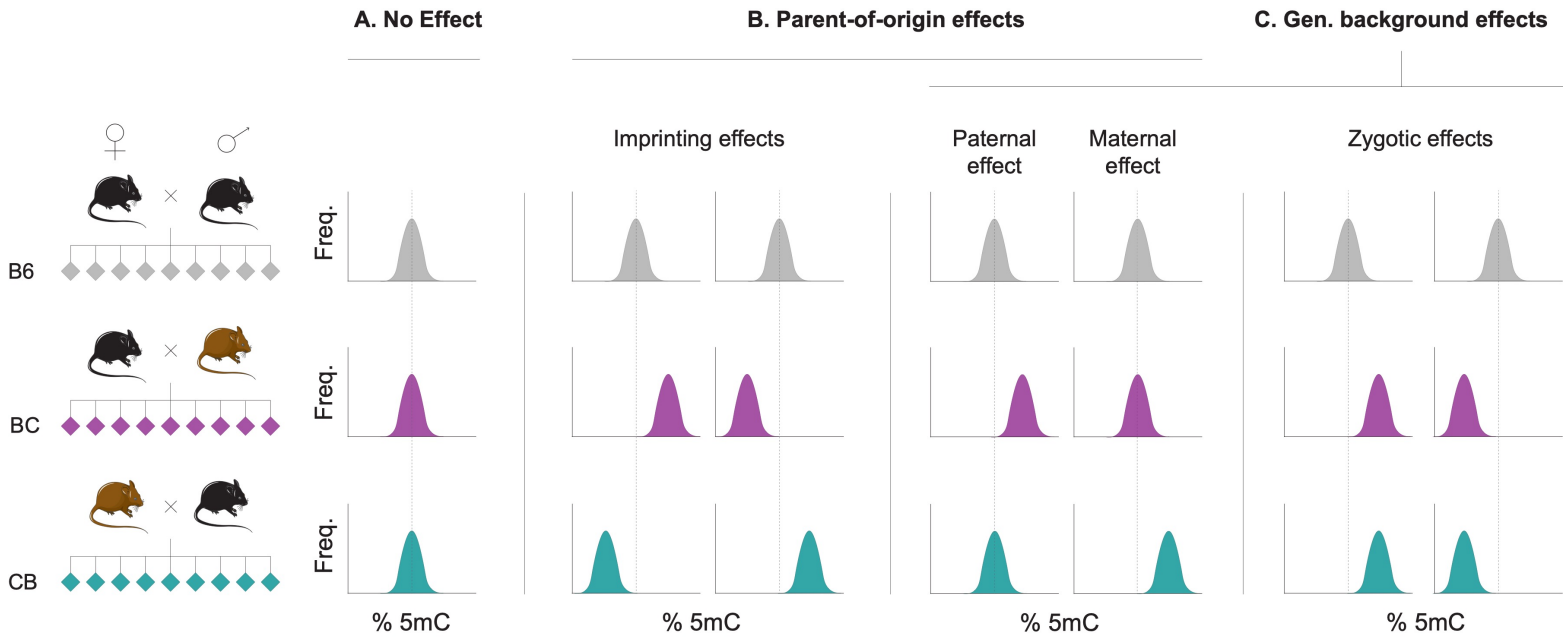


Figure 5.1: Study design and potential outcomes. BC F1 hybrids (purple diamonds) were generated by breeding B6 females (black) with CAST males (brown). CB F1 hybrids (teal diamonds) were produced with the reciprocal cross of CAST females (brown) and B6 males (black). Hypothetical frequency distributions of VM-IAP methylation levels (% 5mC) are shown for pure B6 (grey) and hybrid (purple and teal) F1 offspring. **A.** Parent-of-origin and genetic background are not contributors to VM-IAP methylation levels if the B6, BC, CB frequency distributions are equivalent. **B.** Parent-of-origin effects can be divided into imprinting effects and paternal/maternal effects. Imprinting effects are caused by an intrinsic difference in methylation state between the maternal and paternal alleles and would result in BC and CB methylation distributions shifted in opposite directions, away from the B6 distribution. **C.** Due to the design of this experiment, paternal or maternal effects detected in the analysis must be dependent on the CAST genetic background of the father or mother, respectively, and are therefore classified as both parent-of-origin and genetic background effects. A paternal effect would shift the BC distribution away from the B6 distribution but the CB and B6 distributions would be equivalent. A maternal effect would show the same result in the opposite direction. The diagram depicts paternal and maternal effects that result in an increase in methylation but a decrease in methylation is also possible. Zygotic genetic background effects caused by the introduction of CAST DNA regardless of parental origin would lead to BC and CB methylation distributions shifted in the same direction.

methylation levels form skewed bell-shaped frequency distributions.

5.2.1.2 *Parent-of-origin effects at VM-IAPs are driven by maternal genetic background*

Of the 12 VM-IAPs examined, eight (VM-IAP$_{Mbnl1}$, VM-IAP$_{Tfpi}$, VM-IAP$_{Bmf}$, VM-IAP$_{Rnf157}$, VM-IAP$_{Eps8l1}$, VM-IAP$_{Diap3}$, VM-IAP$_{Rab6b}$, and VM-IAP$_{Sema6d}$) showed significant differences between BC and CB methylation levels, indicating that most VM-IAPs are subject to parent-of-origin effects (**Figure 5.2A** and **C**; **Figure 5.3A** and **C**; Appendix A, **Figure A.4** and **Figure A.6**). For all loci showing this effect, CB distributions were right-shifted compared to BC distributions – in other words, paternal transmission of the VM-IAP following the breeding of CAST females to B6 males (CB) resulted in an increase in VM-IAP methylation compared to maternal transmission of the VM-IAP following the breeding of B6 females to CAST males (BC).

If the parent-of-origin effects at these loci were imprinted and hence a reflection of intrinsic differences in methylation between the parental alleles in the early embryo, the BC distributions would be left-shifted compared to the B6 distributions (mirroring the right-shifted CB distributions). This is because the B6 distribution represents the average of the two alleles, so the single allele in the BC and CB distributions would show reciprocal directional shifts. Such counterbalanced hybrid distributions were not observed (**Figure 5.2A** and **C**; **Figure 5.3A** and **C**; Appendix A, **Figure A.4** and **Figure A.6**), indicating that none of the eight VM-IAPs exhibiting parent-of-origin effects are subject to genomic imprinting.

Instead, with the exception of VM-IAP$_{Rab6b}$ and VM-IAP$_{Sema6d}$ (discussed in the next section), the BC distributions were equivalent to the B6 distributions. This indicates that the parent-of-origin effects are driven by maternal genetic background: a paternal CAST genome appears to have little effect on VM-IAP methylation levels, while a maternal CAST genome results in an increase in methylation. This maternal effect is likely due to oocyte-derived factors. These could be methylation-promoting CAST oocyte-derived factor(s) that are not expressed in the B6 oocyte, or methylation-suppressing B6 oocyte-derived factor(s) that are not present in the CAST oocyte.

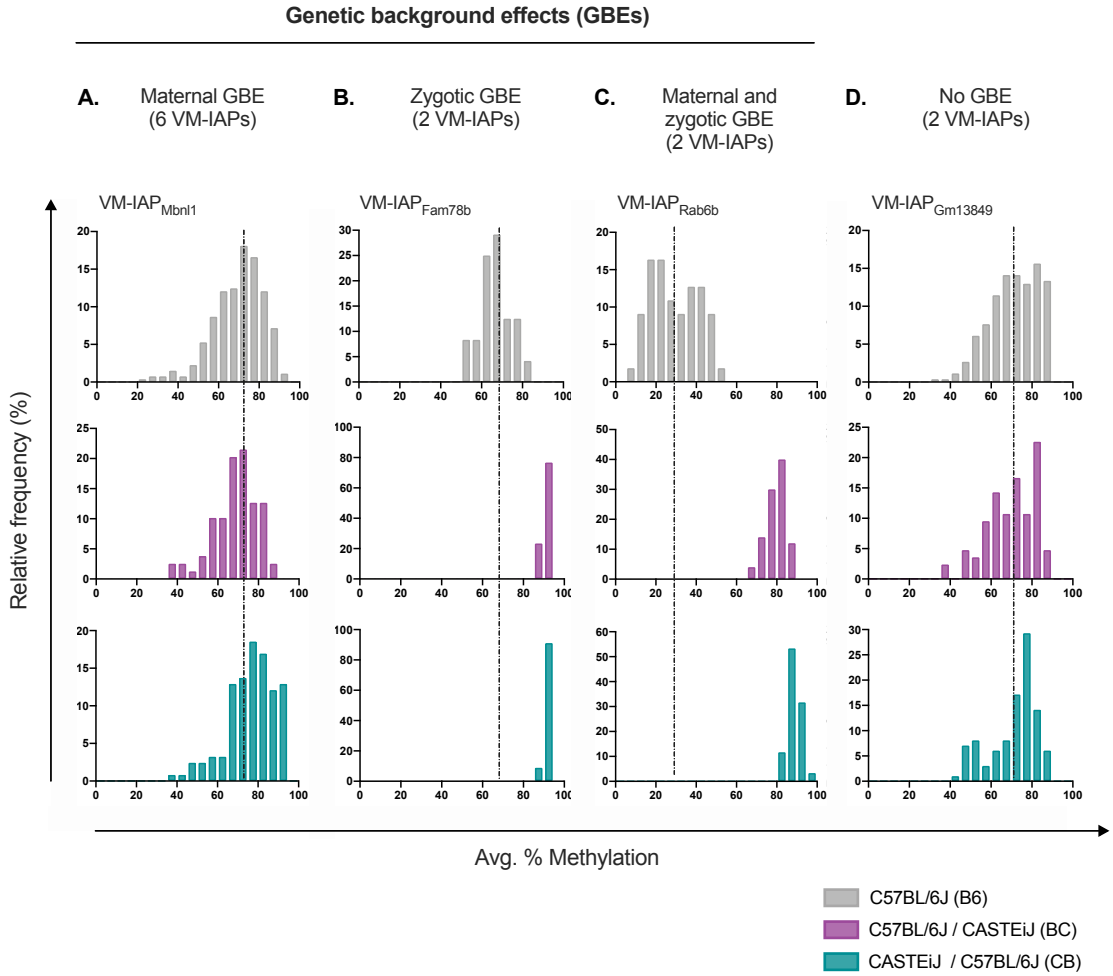


Figure 5.2: Frequency distributions of VM-IAP methylation levels in B6, BC, and CB populations reveal maternal and zygotic genetic background effects (GBEs). A. VM-IAP$_{Mbnl1}$ displays a modest maternal GBE. **B.** VM-IAP$_{Fam78b}$ displays a zygotic GBE. **C.** VM-IAP$_{Rab6b}$ displays both maternal and zygotic GBEs. **D.** VM-IAP$_{Gm13849}$ does not display a GBE. One example is shown per effect type. B6, BC, and CB frequency distributions for all 12 tested VM-IAPs can be found in Appendix A (**Figure A.4**; **Figure A.5**; **Figure A.6**, and **Figure A.7**). Methylation levels were averaged across the distal CpGs of the VM-IAP 5′ LTRs for each individual and classified into 20 bins (bin width = 5 % methylation). Relative frequency was tabulated as a percentage of the population. Dashed vertical lines mark the mean VM-IAP methylation level in B6 mice and are included to facilitate the comparison of BC and CB distributions to B6 distributions.

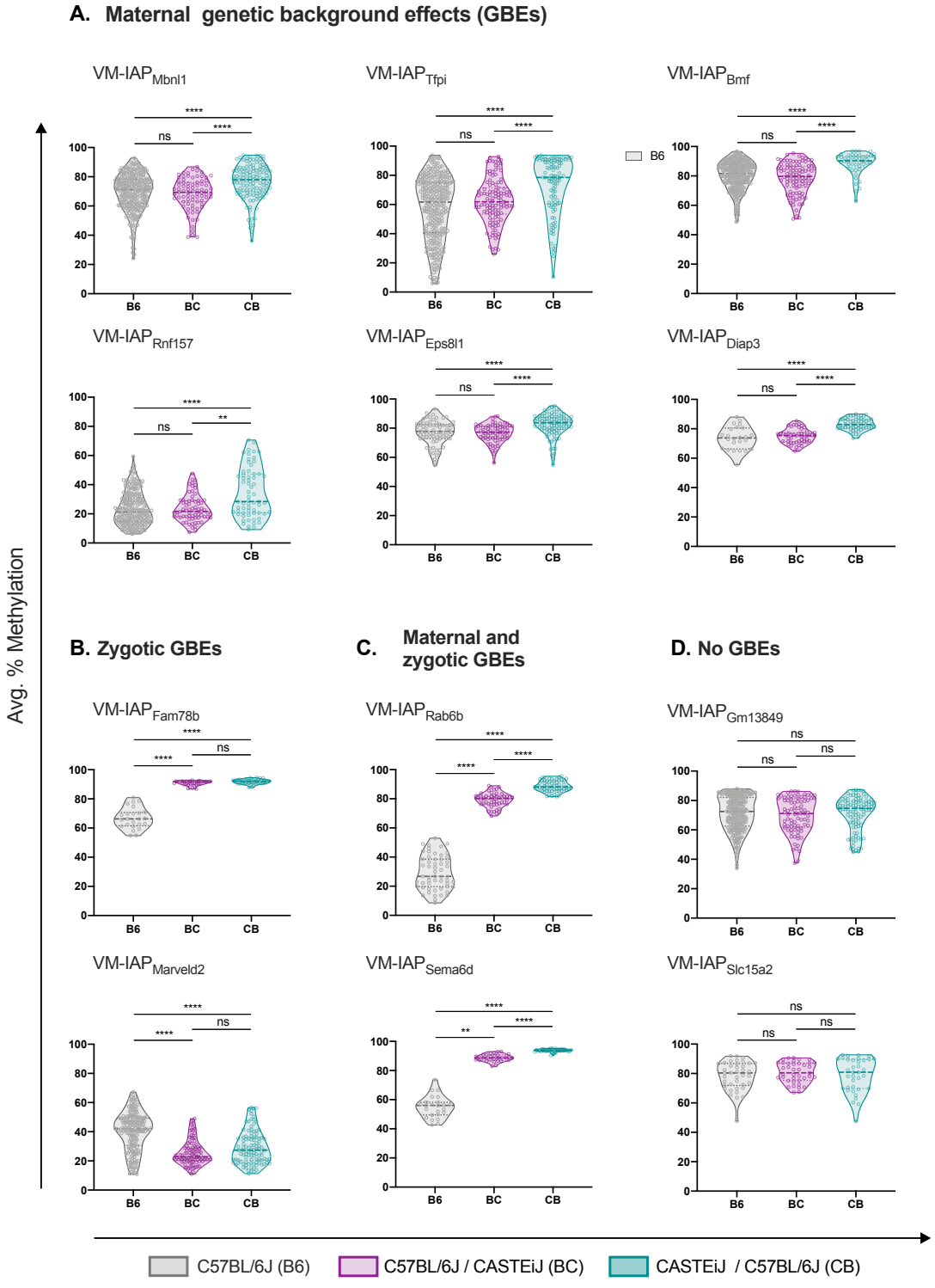


Figure 5.3: VM-IAPs exhibit genetic background effects (GBEs). VM-IAPs are classified based on their susceptibility to maternal GBEs (**A**), zygotic GBEs (**B**), maternal and zygotic GBEs (**C**), or neither (**D**). Violin plots represent the B6 (grey), BC (purple), and CB (teal) F1 offspring methylation distributions. Dotted and dashed lines show the distribution quartiles and median, respectively. Faint hollow circles represent individual-specific methylation levels, averaged across the distal CpGs of the VM-IAP 5′ LTR. B6, BC, and CB distributions were compared for each VM-IAP using the Kruskal-Wallis test followed by Dunn's post hoc multiple comparison test (** $p < 0.005$; **** $p<0.0001$; ns, not significant).

While maternal effects have been reported in a number of contexts, with more than 60 maternal effect genes identified in mammals (reviewed in Condic 2015), it remains possible that the maternal effects observed at VM-IAPs are caused by differences between the CAST and B6 intrauterine environments. Embryo transfer experiments whereby fertilised CAST eggs are transferred into B6 foster dams would be required to assess this possibility. Given that the maternal effects identified in this part of the analysis are strain-specific, the term *maternal genetic background effect* will be used to describe them hereafter to avoid confusion with parent-of-origin effects associated with genomic imprinting.

The equivalent methylation distributions in BC and B6 populations for many of the VM-IAPs shows that hemizygous individuals are capable of recapitulating the full inter-individual methylation ranges observed in homozygous individuals. This suggests that the process leading to stochastic methylation of VM-IAPs occurs independently at each allele.

5.2.1.3 VM-IAP methylation distributions are influenced by zygotic genetic background

The comparison of BC and CB methylation distributions described above exposed maternal genetic background effects at VM-IAPs. In this section, BC and CB distributions are compared to the parental B6 distribution to identify zygotic genetic background effects, defined here as changes caused by the introduction of a CAST haploid genome regardless of parental origin. The CAST distribution is not considered because the 12 VM-IAPs investigated in this study are absent from CAST mice.

Four VM-IAPs (VM-IAP$_{Rab6b}$, VM-IAP$_{Sema6d}$, VM-IAP$_{Fam78b}$, and VM-IAP$_{Marveld2}$) showed significant shifts in methylation when either BC or CB F1 hybrids were compared to B6 individuals (**Figure 5.2B** and **C**; **Figure 5.3B** and **C**; Appendix A, **Figure A.5** and **Figure A.6**). While VM-IAP$_{Rab6b}$ and VM-IAP$_{Sema6d}$ also displayed maternal genetic background effects, their susceptibility to zygotic genetic background effects was more pronounced (**Figure 5.3C**). VM-IAP$_{Marveld2}$ was alone in showing a reduction in methylation levels in F1 hybrids compared to B6 individuals. The other three (VM-IAP$_{Rab6b}$, VM-IAP$_{Sema6d}$, and VM-IAP$_{Fam78b}$) were hypermethylated in BC and CB hybrids, suggesting that in this genomic context

they are targeted for repression like the majority of IAPs in the mouse genome. Of note, VM-IAP$_{Rab6b}$, VM-IAP$_{Sema6d}$, and VM-IAP$_{Fam78b}$ are all solo LTRs of the IAPLTR2_Mm subtype while the nine other VM-IAPs analysed in this chapter have internal protein-coding sequences flanked by two LTRs (**Table 2.1**).

No significant differences were observed between B6, BC, and CB methylation distributions for VM-IAP$_{Gm13849}$ and VM-IAP$_{Slc15a2}$ (**Figure 5.2D**; **Figure 5.3D**; Appendix A, **Figure A.7**). This may be an intrinsic characteristic of these two loci. Alternatively, it is possible that sensitivity to genetic background effects varies across VM-IAPs depending on the mouse strains used in the experiment.

Table 5.1: Maternal and zygotic genetic background effects at VM-IAPs.

VM-IAP	Maternal GBE*	Zygotic GBE
VM-IAP$_{Mbnl1}$	X	
VM-IAP$_{Tfpi}$	X	
VM-IAP$_{Bmf}$	X	
VM-IAP$_{Rnf157}$	X	
VM-IAP$_{Eps8l1}$	X	
VM-IAP$_{Diap3}$	X	
VM-IAP$_{Fam78b}$		X
VM-IAP$_{Marveld2}$		X
VM-IAP$_{Rab6b}$	X	X
VM-IAP$_{Sema6d}$	X	X
VM-IAP$_{Gm13849}$		
VM-IAP$_{Slc15a2}$		

*GBE = genetic background effect

5.2.2 Manipulation of VM-IAP methylation states via genetic backcrossing

To determine whether the induced hypermethylation at VM-IAP$_{Rab6b}$, VM-IAP$_{Sema6d}$, and VM-IAP$_{Fam78b}$ in F1 hybrids is reversible, BC and CB individuals were backcrossed to B6 to produce N2 offspring. Methylation variation was re-established in the N2 generation at all three VM-IAPs irrespective of parental origin, indicating that the F1 hybrid methylation state is not a permanent one (**Figure 5.4**). However, the inter-individual methylation ranges in the N2

generation did not reflect those observed in B6 mice, suggesting that the CAST genomic content (25%) of these mice is sufficient to disrupt VM-IAP methylation ranges. A superficial assessment of the N2 generation data indicates that backcrossing F1 hybrids to B6 mice causes a widening of the VM-IAP$_{Rab6b}$ inter-individual methylation range, with the potential for progressive reduction upon successive rounds of B6 backcrossing until fully recapitulating the original B6 range.

Interestingly, combining all N2 individuals on a single graph revealed a bi-modal segregation of VM-IAP$_{Rab6b}$ methylation levels (**Figure 5.4A**, right panel). This potentially reflects a return to the B6 methylation state in half of the N2 mice and may suggest genetic segregation of a modifier of metastability. To explore this possibility, a methylation level of 60% was selected as a cut-off value to distinguish B6 from F1 hybrid methylation levels. 60% was deemed the most appropriate threshold because the VM-IAP$_{Rab6b}$ methylation frequency distribution in the B6 population never exceeds 60% (**Figure 5.2C**). This cut-off value was subsequently used to assign each N2 individual with one of two "methylation phenotypes": lowly methylated (reflecting the B6 methylation state) or highly methylated (reflecting the F1 hybrid state). This resulted in an approximate 50:50 split in methylation phenotypes.

The inheritance pattern observed at VM-IAP$_{Rab6b}$ might suggest a hemizygosity effect, whereby VM-IAP$_{Rab6b}$ is lowly methylated if both of its allelic copies are present and highly methylated if only one is. However, this was not found to be the case. PCR primers were designed to target the regions up- and downstream of the VM-IAP$_{Rab6b}$ solo LTR, producing a ~700bp fragment when amplification occurs from the B6 allele and a ~100bp fragment when amplification occurs from the CAST allele lacking the IAP element. Therefore, mice that were homozygous for the IAP only amplified the ~700 bp fragment and mice that were hemizygous for the IAP amplified both fragments. Hemizygous individuals were represented in both the highly and lowly methylated groups, indicating that VM-IAP hemizygosity is not driving the segregation in methylation states (**Figure 5.5**).

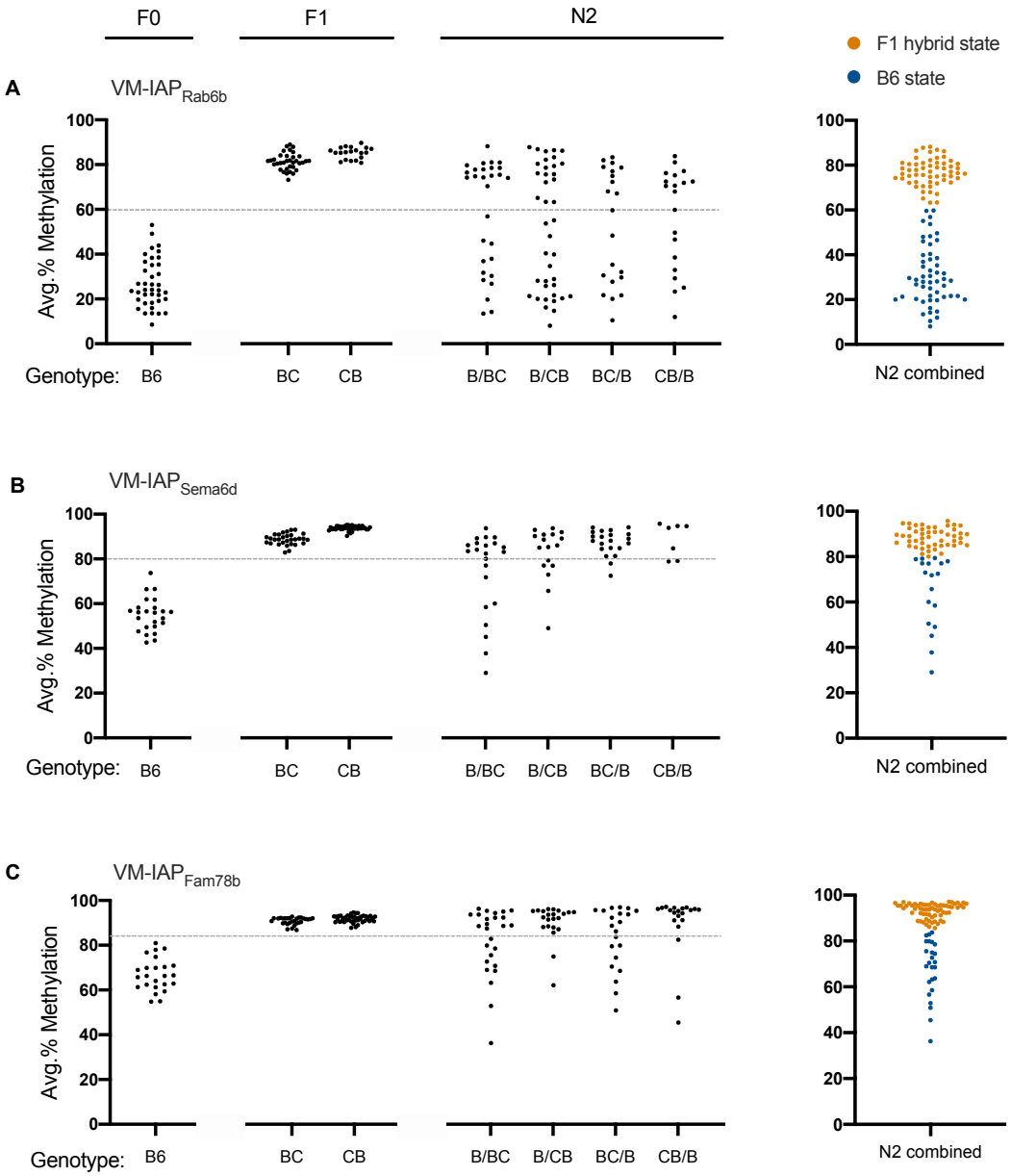


Figure 5.4: Recovery of variable epigenetic states via genetic backcrossing. Methylation variation is reacquired at VM-IAP$_{Rab6b}$ (**A**), VM-IAP$_{Sema6d}$ (**B**), and VM-IAP$_{Fam78b}$ (**C**) in the N2 generation following B6 backcrossing. Methylation levels were averaged across the distal CpGs of the VM-IAP 5′ LTRs for each individual. Genotype notation for N2 individuals indicates the maternal genotype followed by the paternal genotype, separated by a slash. Right panels show N2 individuals combined into one graph and colour-coded based on whether their methylation level reflects the B6 (navy) or F1 hybrid (orange) methylated state. Dashed lines represent the cut-off value used to carry out this methylation phenotyping for each VM-IAP. The N2 methylation data were generated in collaboration with Jessica Elmer with assistance from Jesse Loilargosain.

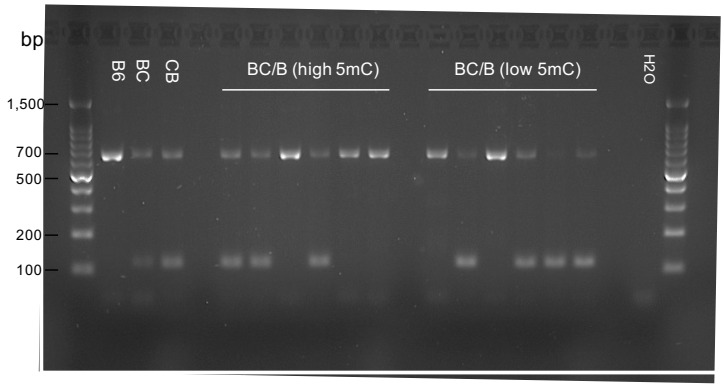


Figure 5.5: Segregation of VM-IAP$_{Rab6b}$ methylation states in N2 mice is independent of VM-IAP$_{Rab6b}$ copy number. Six N2 individuals of each methylation phenotype were genotyped via PCR using primers targeting the VM-IAP$_{Rab6b}$ flanking regions, amplifying a ~700bp B6 fragment and a ~100bp CAST fragment. PCR product sizes were assessed via agarose gel electrophoresis. Individuals only amplifying the larger fragment are homozygous for VM-IAP$_{Rab6b}$ and individuals amplifying both fragments are hemizygous for VM-IAP$_{Rab6b}$. The B6, BC, and CB DNA samples are included as controls.

An alternative explanation for the 1:1 ratio of VM-IAP$_{Rab6b}$ methylation states in N2 mice is the Mendelian inheritance of a dominant *trans*-acting strain-specific modifier (*Ssm*) gene in the CAST genome that is absent from the B6 genome. In this scenario, one copy of the hypothetical modifier gene would be present in all F1 hybrid offspring, the product of which would act as a dominant repressor of VM-IAP$_{Rab6b}$. Upon backcrossing to B6, half of the N2 offspring would inherit the CAST-derived allele, resulting in a bimodal distribution of methylation states. A Punnett square illustrating this inheritance pattern is shown in **Figure 5.6A**, using the generic *Ssm$^{VM\text{-}IAP}$* notation with *B* and *C* superscripts to denote the B6 and CAST alleles, respectively. This nomenclature highlights the possibility of there being a version of the modifier gene in the B6 genome coding for a protein that is functionally distinct from the CAST allele. A chi-squared test confirmed that the VM-IAP$_{Rab6b}$ N2 methylation data follow the expected 1:1 ratio (**Table 5.2**).

The ability to differentiate between the two VM-IAP$_{Rab6b}$ methylation phenotypes is a fortunate consequence of the lowly methylated variable state of this VM-IAP in a pure B6 genetic background. The N2 methylation values at VM-

IAP$_{Sema6d}$ and VM-IAP$_{Fam78b}$ are more difficult to interpret because these loci are more highly methylated in B6 mice. Nonetheless, the same approach was used to analyse the N2 methylation data at these two VM-IAPs: the methylation value that best separated B6 from F1 hybrid mice (80% for VM-IAP$_{Sema6d}$ and 85% for VM-IAP$_{Fam78b}$) was used as a cut-off to classify N2 mice into two cohorts (**Figure 5.4B** and **C**).

Chi-squared analyses are implemented to assess the goodness of fit between observed and theoretically expected data. Chi-squared tests using an expected ratio of 1:1 showed that N2 methylation states at VM-IAP$_{Sema6d}$ and VM-IAP$_{Fam78b}$ are statistically unlikely to be sampled from a 1:1 population distribution (**Table 5.2**). When an expected ratio of 3:1 was applied, the N2 data for both VM-IAPs appeared to fit a 3:1 population distribution (**Table 5.2**). The Punnet square in **Figure 5.6B** depicts the expected 1:1:1:1 genotypic and 3:1 phenotypic outcomes of crossing an F1 hybrid mouse carrying two unlinked CAST-derived modifiers genes (*SsmX*$_{VM\text{-}IAP}^{C}$ and *SsmY*$_{VM\text{-}IAP}^{C}$) with a B6 mouse carrying neither. It is therefore possible that VM-IAP$_{Sema6d}$ and VM-IAP$_{Fam78b}$ are targeted by two dominant SSMs, leading to the reestablishment of methylation variation in 25% of N2 backcrossed offspring. The reacquisition of variable methylation levels at VM-IAP$_{Rab6b}$, VM-IAP$_{Sema6d}$, and VM-IAP$_{Fam78b}$ after a single generation of backcrossing is indicative of limited redundancy of VM-IAP modifiers.

To explore whether VM-IAP$_{Rab6b}$, VM-IAP$_{Sema6d}$, and VM-IAP$_{Fam78b}$ are modulated by overlapping SSMs, a heatmap was generated comparing the methylation phenotypes of N2 offspring across the three loci (**Figure 5.6C**). VM-IAP$_{Sema6d}$ and VM-IAP$_{Fam78b}$ did not show concordance across individuals, suggesting that different SSMs are involved in establishing methylation patterns at these two VM-IAPs (**Figure 5.6C**). In contrast, individuals that were lowly methylated at VM-IAP$_{Sema6d}$ (i.e. reflecting the B6 methylation state) were also lowly methylated at VM-IAP$_{Rab6b}$, indicating that the hypothetical SSM$_{Rab6b}$ may be one of the two SSMs potentially targeting VM-IAP$_{Sema6d}$. Following this logic, individuals lowly methylated at VM-IAP$_{Rab6b}$ and highly methylated at VM-IAP$_{Sema6d}$ would be carriers of the non-shared *Ssm* gene, making up a useful subgroup to consider in future mapping experiments. There was one mouse that was lowly methylated at VM-IAP$_{Sema6d}$ and highly methylated at VM-IAP$_{Rab6b}$ but this individual was likely incorrectly phenotyped: its VM-IAP$_{Sema6d}$ methylation

level was 79.1%, just one percentage point under the arguably crude 80% cut-off. This intra-individual comparative analysis demonstrates that not all VM-IAPs are targeted by the same modifier(s) and supports the prospect of shared modifiers across a subset of VM-IAPs.

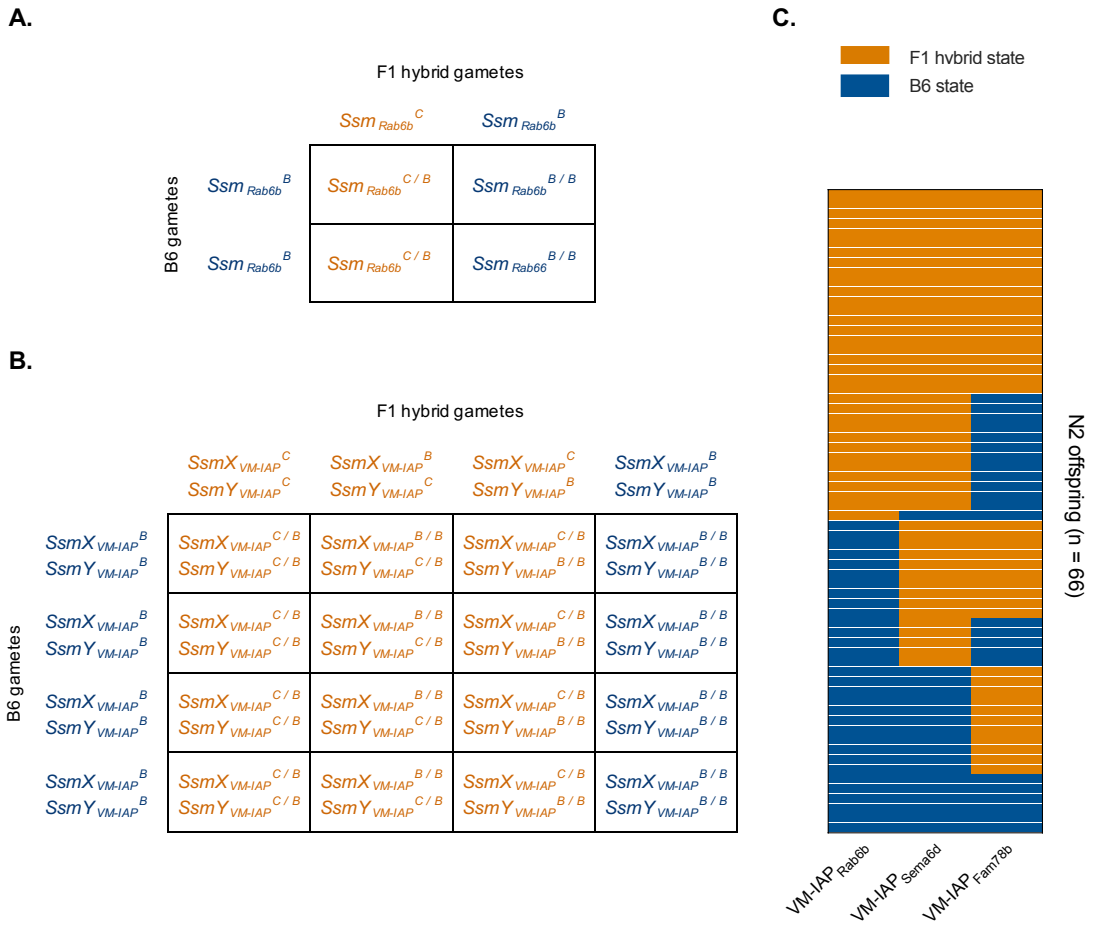


Figure 5.6: VM-IAP$_{Rab6b}$ and VM-IAP$_{Sema6d}$ likely share a dominant CAST modifier. A. Punnett square analysis depicting the expected 1:1 phenotypic outcome of crossing a hypothetical *Ssm*$_{VM\text{-}IAP^{C/B}}$ heterozygote (orange) with a homozygous *Ssm*$_{VM\text{-}IAP^{B/B}}$ B6 mouse (navy). **B.** Punnett square analysis depicting the expected 3:1 phenotypic outcome of crossing a hypothetical individual heterozygote for two *Ssm*s (*SsmX*$_{VM\text{-}IAP^{C/B}}$ and *SsmY*$_{VM\text{-}IAP^{C/B}}$, orange) with a homozygous *SsmX*$_{VM\text{-}IAP^{B/B}}$, *SsmY*$_{VM\text{-}IAP^{B/B}}$ B6 mouse (navy). The *B* and *C* superscripts differentiate the hypothetical B6 and CAST alleles, respectively. **C.** Heatmap comparing the methylation phenotypes of 66 N2 backcrossed offspring across VM-IAP$_{Rab6b}$, VM-IAP$_{Sema6d}$, and VM-IAP$_{Fam78b}$.

Table 5.2: Chi-squared analysis on N2 data.

	Methylation state		Totals	p-value[3]
	High	Low (variable)		
VM-IAP$_{Rab6b}$				
Observed	57	53	110	
Expected (1 SSM)[1]	55	55	110	0.703
Expected (2 SSMs)[2]	82.5	27.5	110	1.97E-08
VM-IAP$_{Sema6d}$				
Observed	49	18	67	
Expected (1 SSM)[1]	33.5	33.5	67	1.52E-04
Expected (2 SSMs)[2]	50.25	16.75	67	0.724
VM-IAP$_{Fam78b}$				
Observed	64	25	89	
Expected (1 SSM)[1]	44.5	44.5	89	3.57E-05
Expected (2 SSMs)[2]	66.75	22.25	89	0.501

[1] Expected 1:1 inheritance pattern of a dominant autosomal modifier.
[2] Expected 3:1 inheritance pattern of two dominant unlinked autosomal modifiers.
[3] Likelihood that the values were sampled from the expected distribution, calculated using the CHISQ.TEST function in excel.

As mentioned above, VM-IAP$_{Rab6b}$, VM-IAP$_{Sema6d}$, and VM-IAP$_{Fam78b}$ are solo LTRs of the IAPLTR2_Mm subclass and do not possess internal protein coding regions. This limits the binding site of their hypothetical SSMs to their ~500bp LTR sequence (under the assumption that these loci are targeted via the direct or indirect binding of their DNA sequences). Given that VM-IAP$_{Rab6b}$ and VM-IAP$_{Sema6d}$ may be repressed by the same SSM, their sequences were expected to be more similar to each other than to the sequence of VM-IAP$_{Fam78b}$. Indeed, while all three VM-IAPs show approximately 70% identity with the IAPLTR2_Mm consensus sequence, VM-IAP$_{Rab6b}$ shares 92.12% identity with VM-IAP$_{Sema6d}$ and 83% identity with VM-IAP$_{Fam78b}$. VM-IAP$_{Sema6d}$ and VM-IAP$_{Fam78b}$ share 84.68% identity (**Table 5.3**). This analysis indicates that bulk sequence identity may be useful in identifying SMM targets, as will be further emphasised in section 5.2.3.2. The sequence alignment depicted in **Figure 5.7**

Table 5.3: Percent identity matrix[1].

	IAPLTR2_Mm[2]	VM-IAP$_{Rab6b}$[3]	VM-IAP$_{Sema6d}$[3]	VM-IAP$_{Fam78b}$[3]
IAPLTR2_Mm	100			
VM-IAP$_{Rab6b}$	74.13	100		
VM-IAP$_{Sema6d}$	73.33	92.12	100	
VM-IAP$_{Fam78b}$	72.92	83.00	84.68	100

[1]Percent identity matrix created by Clustal 2.1
[2]IAPLTR2_Mm consensus sequence extracted from Dfam.org
[3]VM-IAP sequences extracted from the UCSC genome browser

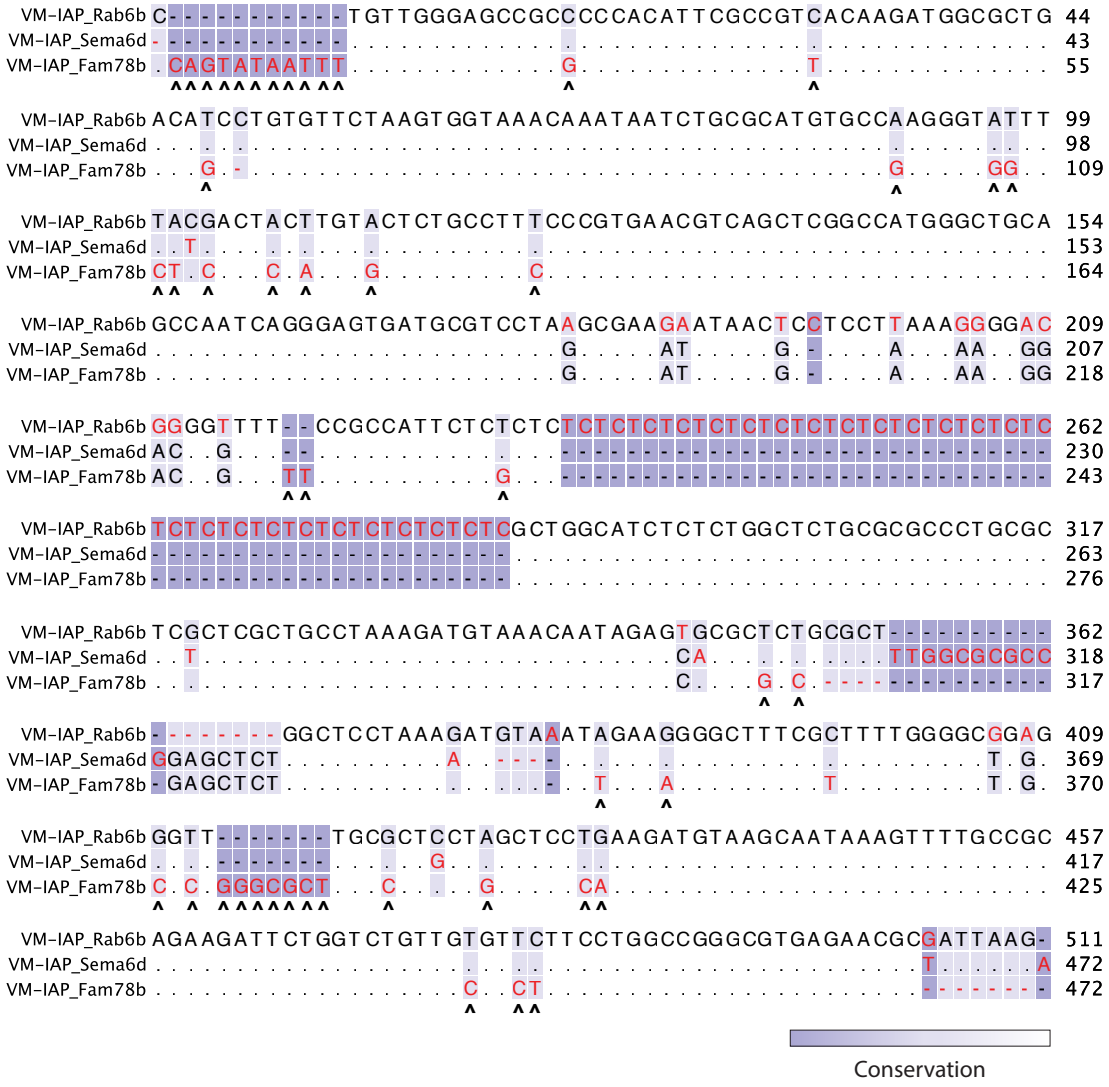


Figure 5.7: Sequence alignment of VM-IAP$_{Rab6b}$, VM-IAP$_{Sema6d}$, and VM-IAP$_{Fam78b}$. The VM-IAP sequences were extracted from the UCSC genome browser. The VM-IAP$_{Sema6d}$ and VM-IAP$_{Fam78b}$ sequences were reverse-complemented prior to generating the alignment. Bases are colour-coded based on conservation and dashes indicate lack of sequence. Arrowheads designate regions that are shared between VM-IAP$_{Rab6b}$ and VM-IAP$_{Sema6d}$ but not VM-IAP$_{Fam78b}$. The alignment was generated with CLC Sequence Viewer 6.

draws attention to the regions shared by VM-IAP$_{Rab6b}$ and VM-IAP$_{Sema6d}$, but not VM-IAP$_{Fam78b}$ (shown with arrowheads). These regions are candidate SSM binding sites.

5.2.3 Establishing VM-IAP methylation as an effective phenotypic readout for the genetic mapping of SSMs

5.2.3.1 VM-IAP$_{Rab6b}$ is targeted for methylation by a single CAST-derived SSM.

The previous section demonstrated that metastability can be reacquired at VM-IAPs via genetic backcrossing and outlined how VM-IAP methylation levels might function as molecular phenotypes associated with SSMs. Punnett square analyses using hypothetical *Ssm* genes showed how this experimental design could be used to predict the number of SSMs targeting a given VM-IAP. However, it is acknowledged that the conclusions reached above are just predictions at this point. This is particularly true for the inferences associated with VM-IAP$_{Sema6d}$ and VM-IAP$_{Fam78b}$ because the methylation values chosen as phenotypic cut-offs for these alleles are not as clear as the one specified for VM-IAP$_{Rab6b}$. In addition, it is likely that the maternal genetic background effects associated with VM-IAP$_{Sema6d}$ had a partial influence on the results of the chi-squared tests. Therefore, due to the unambiguity of its methylation phenotype and the apparent simplicity of its inheritance pattern, VM-IAP$_{Rab6b}$ was selected as the focus of a proof-of-principle experiment aimed at establishing the B6/CAST hybrid system as an effective tool for the genetic mapping and functional characterisation of VM-IAP SSMs.

To test the hypothesis that N2 mice exhibiting hypermethylation of VM-IAP$_{Rab6b}$ are carriers of a CAST-derived *Ssm* gene, an additional round of B6 backcrossing was conducted. Highly and lowly methylated N2 males were bred with B6 females and VM-IAP$_{Rab6b}$ methylation levels were quantified in the N3 generation. N3 offspring generated from highly methylated males recreated the bimodal distribution observed in N2 offspring, producing a 1:1 ratio of high-to-low methylation levels. N3 offspring generated from lowly methylated N2 males were all lowly methylated and displayed the range originally observed in inbred

B6 mice (**Figure 5.8A**). This result, summarised in pedigree form in **Figure 5.8B**, indicates that a single CAST-derived modifier, $SSM_{Rab6b}{}^{C}$, is in fact responsible for the methylation of VM-IAP_{Rab6b} in hybrids. Experiments designed to determine the genomic location of $Ssm_{Rab6b}{}^{C}$ are ongoing.

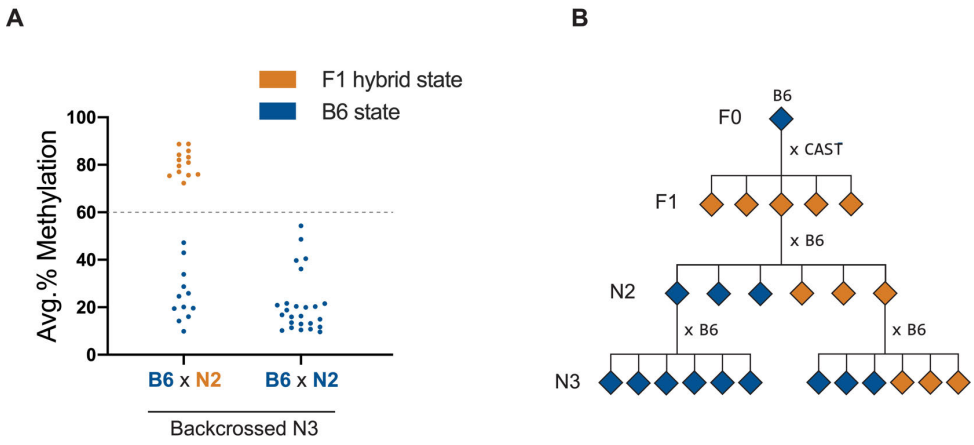


Figure 5.8: VM-IAP_{Rab6b} is targeted for methylation by a single CAST-derived dominant SSM. A. N2 males of each methylation phenotype were backcrossed to B6 females and VM-IAP_{Rab6b} 5′ LTR methylation levels were quantified in ear samples from N3 offspring. Three N3 litters are represented per N2 phenotype, generated from separate parents. Data points represent average methylation level across CpGs for each individual, colour-coded based on whether their methylation level reflects the B6 (navy) or F1 hybrid (orange) methylated state. The X axis labels adhere to the same colour code. The dashed line represents the 60% cut-off value used to determine VM-IAP_{Rab6b} methylation phenotypes. **B.** Pedigree illustrating the inheritance patterns of VM-IAP_{Rab6b} methylation states.

It is possible that the stochastic methylation of VM-IAPs is symptomatic of recent IAP integration, representing a snapshot in evolutionary time as the host genome adapts to repress new insertions. The efficient silencing of VM-IAP_{Rab6b} by $SSM_{Rab6b}{}^{C}$ despite the absence of VM-IAP_{Rab6b} from the CAST genome seems at odds with this hypothesis. To investigate whether VM-IAP_{Rab6b} was present in the CAST genome in evolutionary past, an alignment was generated to compare the B6 VM-IAP_{Rab6b} insertion site and its homologous CAST sequence. The 6bp target site duplication (TSD) characteristic of retrotransposition events was observed on either side of VM-IAP_{Rab6b} in B6, but there were no TSDs in the CAST sequence (**Figure 5.9**). The rest of the integration site shows near-perfect

alignment between the two strains, arguing that an IAP insertion has never been present at this locus in the CAST genome.

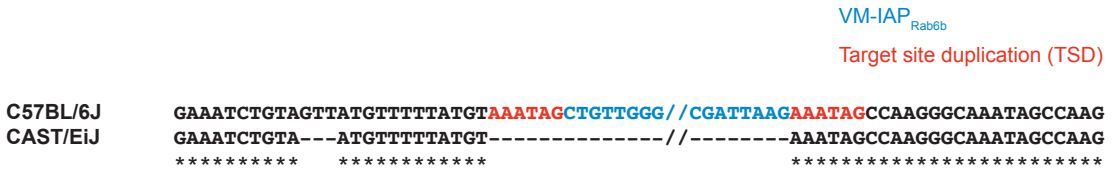


Figure 5.9: VM-IAP$_{Rab6b}$ was not present in the CAST genome in evolutionary past. A B6 and CAST sequence alignment of the VM-IAP$_{Rab6b}$ insertion site was generated using CLUSTAL Omega. B6 and CAST DNA sequences were extracted from the GRCm38/mm10 and CAST_EiJ_v1 assemblies, respectively, accessed through the UCSC genome browser. The VM-IAP$_{Rab6b}$ sequence is shown in blue and target site duplication (TSD) sequences are shown in red. Asterisks indicate residue conservation. The majority of the repeat element was omitted from the B6 sequence; nothing was omitted from the CAST sequence.

5.2.3.2 Other VM-IAPs are targeted for repression by SSM$_{Rab6b}^{C}$

The most parsimonious mechanism by which SSM$_{Rab6b}^{C}$ induces VM-IAP$_{Rab6b}$ hypermethylation is via the recognition of an LTR binding site. In view of the repetitive nature of IAPs, it would make sense for other IAPs that are highly similar in sequence to VM-IAP$_{Rab6b}$ to be methylated in the presence of SSM$_{Rab6b}^{C}$, and variably methylated in its absence. Using the BLAT alignment tool (Kent 2002), the B6 reference genome was queried for regions of high sequence similarity to VM-IAP$_{Rab6b}$. This resulted in a list of 235 LTRs, many of which were duplicates in that they represented the 5′ and 3′ LTRs of the same IAP element. To determine whether any of these LTRs were associated with variable methylation in B6 mice, the list of LTR coordinates was intersected with VM-IAP coordinates using BedTools software (Quinlan & Hall 2010). In addition to VM-IAP$_{Rab6b}$, six VM-IAPs overlapped both coordinate lists, including VM-IAP$_{Tmprss11d}$, VM-IAP$_{Ect2l}$, VM-IAP$_{Pink1,}$ VM-IAP$_{Trbv31}$, VM-IAP$_{Gm20110}$, and VM-IAP$_{Rps12}$ (**Table 5.4**). All of them are solo LTRs of the IAPLTR2_Mm subclass and none had been selected for detailed analysis at the onset of this thesis. Considering there are more than 15,000 LTRs and fewer than 100 VM-IAPs in

the B6 genome, this number represents a significant enrichment in VM-IAPs and suggests that DNA sequence is a major contributor to VM-IAP metastability.

Table 5.4: List of VM-IAPs represented in VM-IAP$_{Rab6b}$ BLAT hits.

VM-IAP	Score[1]	Identity (%)	mm10 coordinates	Structure	IAP subtype
VM-IAP$_{Rab6b}$	511	100.0	chr9:103108601-103109111	solo LTR	IAPLTR2_Mm
VM-IAP$_{Tmprss11d}$	436	95.2	chr5:86318274-86318748	solo LTR	IAPLTR2_Mm
VM-IAP$_{Ect2l}$	431	96.0	chr10:18139121-18139589	solo LTR	IAPLTR2_Mm
VM-IAP$_{Pink1}$	428	94.7	chr4:138327386-138327856	solo LTR	IAPLTR2_Mm
VM-IAP$_{Trbv31}$	409	94.5	chr6:41563008-41563472	solo LTR	IAPLTR2_Mm
VM-IAP$_{Gm20110}$	386	93.6	chr10:99599559-99600037	solo LTR	IAPLTR2_Mm
VM-IAP$_{Rps12}$	353	94.6	chr10:23726341-23726812	solo LTR	IAPLTR2_Mm

[1]The BLAT score reports the number of matching nucleotides with penalty for mismatches and gaps.

VM-IAP$_{Pink1}$ and VM-IAP$_{Ect2l}$ were arbitrarily selected as initial candidates to test whether the VM-IAPs identified in this analysis are also SSM$_{Rab6b^C}$ targets. Both loci showed variable methylation levels in B6 mice and became hypermethylated in BC and CB F1 hybrids, as observed for VM-IAP$_{Rab6b}$, VM-IAP$_{Sema6d}$, and VM-IAP$_{Fam78b}$ (**Figure 5.10A** and **B**). Next, VM-IAP$_{Pink1}$ and VM-IAP$_{Ect2l}$ methylation levels were quantified in N3 mice born to an N2 heterozygous carrier of the *Ssm$_{Rab6b^C}$* allele (i.e. highly methylated at VM-IAP$_{Rab6b}$). As done previously, a locus-specific methylation cut-off value was specified for each VM-IAP and subsequently used to assign a methylation phenotype to N3 individuals (**Figure 5.10A** and **B**). A cross-locus comparison revealed that the five individuals displaying hypermethylation at VM-IAP$_{Rab6b}$ were also hypermethylated at VM-IAP$_{Pink1}$ and VM-IAP$_{Ect2l}$, while the rest were low- and variably methylated at all three VM-IAPs. This finding demonstrates that VM-IAP$_{Pink1}$ and VM-IAP$_{Ect2l}$ are also targeted by SSM$_{Rab6b^C}$. The other four VM-IAPs with high sequence identity with VM-IAP$_{Rab6b}$ are currently being tested in the same manner.

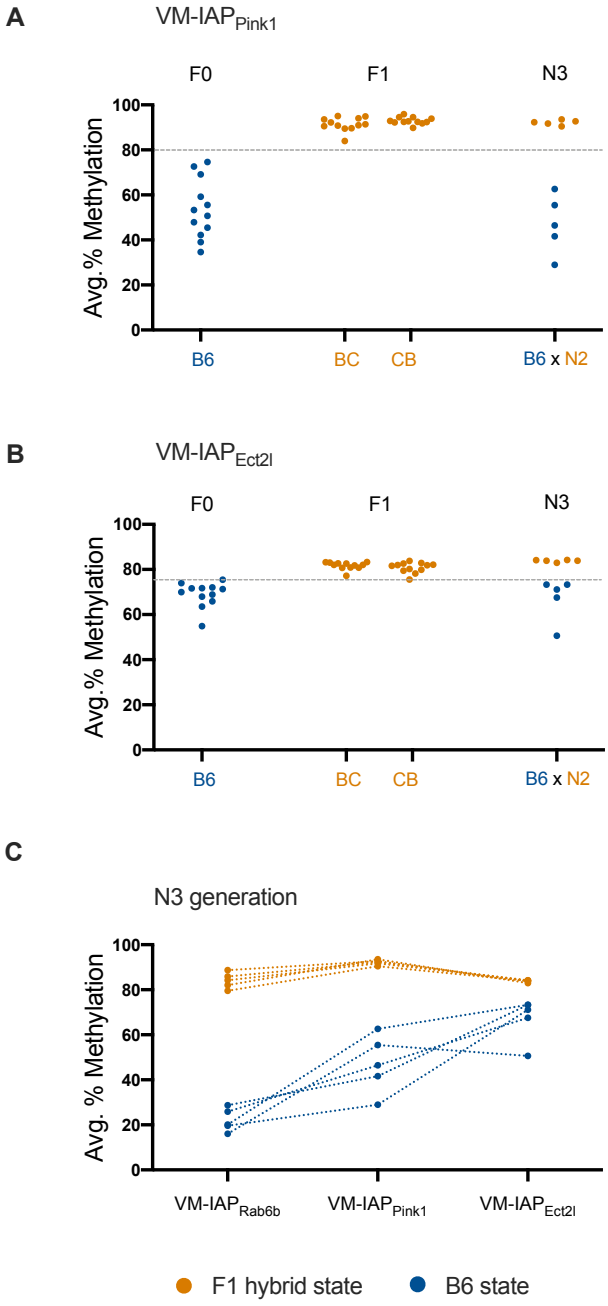


Figure 5.10: VM-IAP$_{Pink1}$ and VM-IAP$_{Ect2l}$ are targeted for repression by SSM$_{Rab6b}$^{C}. VM-IAP$_{Pink1}$ (**A**) and VM-IAP$_{Ect2l}$ (**B**) are variably methylated in B6 individuals, hypermethylated in BC and CB F1 hybrids, and exhibit both methylation phenotypes in the N3 backcrossed generation. **C.** N3 individuals that are highly methylated at VM-IAP$_{Rab6b}$ are also highly methylated at VM-IAP$_{Pink1}$ and VM-IAP$_{Ect2l}$. The same concordance is observed for lowly methylated N3 individuals. Methylation levels were averaged across the distal CpGs of the VM-IAP 5′ LTRs for each individual. Individuals are colour-coded based on whether their methylation level reflects the B6 (navy) or F1 hybrid (orange) methylated state. Dashed grey lines represent the cut-off value distinguishing the B6 methylation state from the F1 hybrid methylation state for each VM-IAP. Orange and navy dotted lines connect the methylation values of the same N3 individual across the three VM-IAPs.

5.3 Discussion

The findings in this chapter highlight the value of harnessing genetic differences between inbred mouse strains to investigate epigenetic phenomena. When comparing BC and CB methylation distributions, most VM-IAPs exhibited parent-of-origin effects that were dependent on maternal genetic background rather than intrinsic allelic differences. The mechanism by which cytoplasmic factor(s) in the CAST oocyte might promote increased VM-IAP methylation in F1 offspring is unresolved. Simplified schematics of potential mechanisms are illustrated in **Figure 5.11**. It is possible that CAST-oocyte-derived gene products target VM-IAPs for repression during the oocyte-to-embryo transition (**Figure 5.11A**). However, this period in early mouse development is characterised by DNA demethylation rather than gain of methylation, at least on a genome-wide scale (Hackett & Surani 2013; Smallwood & Kelsey 2012). This favours the hypothesis that CAST oocyte factors are interfering with TET3-mediated or DNA replication-dependent demethylation of paternally inherited loci (Gu *et al.* 2011), perhaps via the active protection of VM-IAPs (**Figure 5.11B**). This possibility is reminiscent of maternal ZFP57-mediated protection of imprinted loci in the zygote (Li *et al.* 2008; Strogantsev *et al.* 2015). Alternatively, incomplete reprogramming could be the result of divergent evolution, whereby the mechanisms that evolved to efficiently demethylate a CAST paternal genome are not optimised to demethylate a B6 paternal genome (**Figure 5.11C**). While these are important mechanistic considerations, it should be noted that the maternal genetic background effects at VM-IAPs are relatively modest compared to the zygotic genetic background effects: they cause increases in the average methylation level in the population but never abolish inter-individual methylation variability itself. In addition, embryo transfer experiments have not been carried out to rule out intrauterine environment effects.

The possibility of minor parent-of-origin effects at VM-IAPs in the context of a pure B6 genetic background have not been ruled out. This could be tested by introducing a single SNP adjacent to a VM-IAP in an otherwise B6 genome and subsequently carrying out reciprocal crosses between the resulting mice and

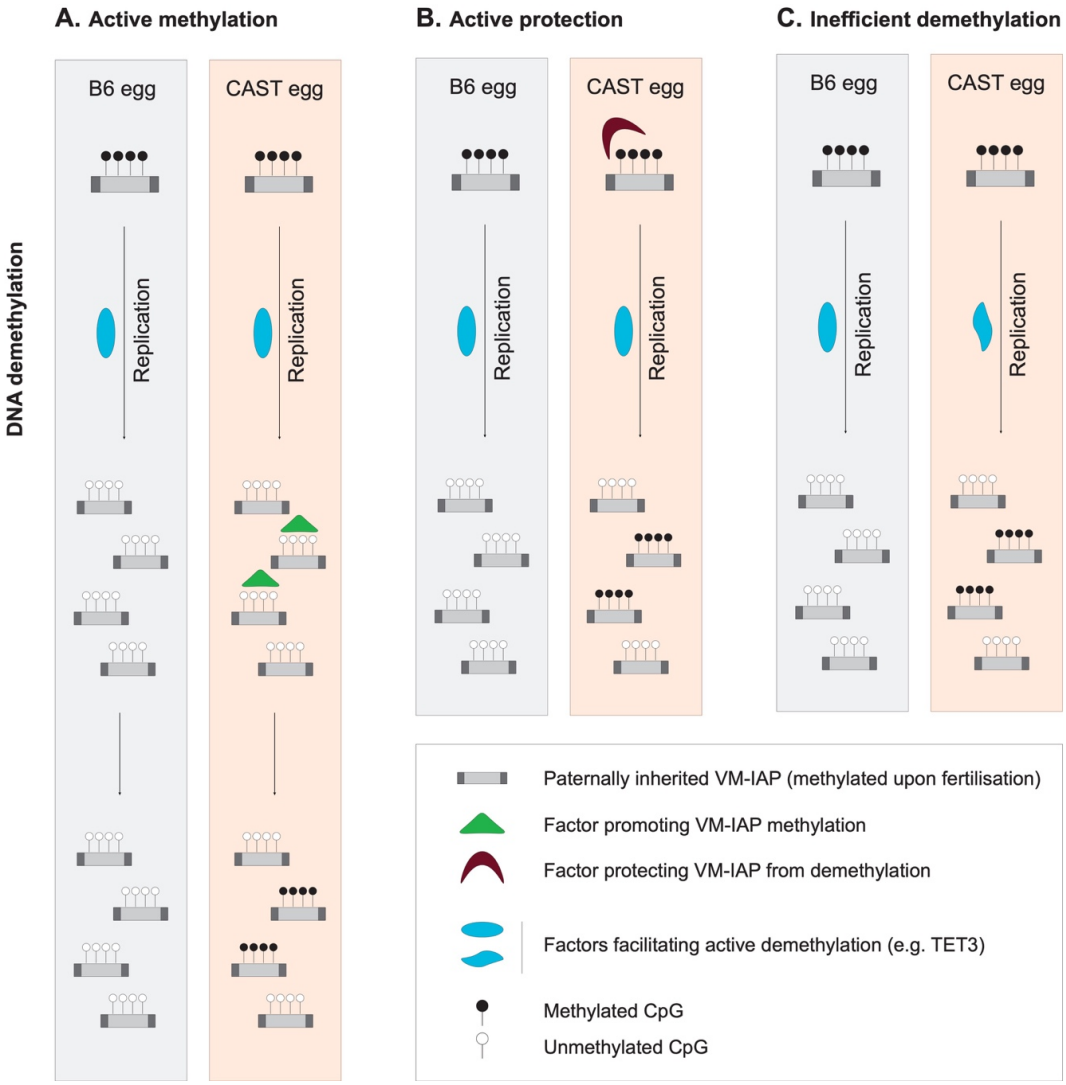


Figure 5.11: Potential mechanisms underlying maternal genetic background effects. Three non-exhaustive mechanisms conceptualise ways in which a CAST oocyte might promote increased VM-IAP methylation levels compared to a B6 oocyte. In all three scenarios, the paternally inherited VM-IAP allele (grey rectangle) enters the egg in a highly methylated state (black circles). In the B6 oocyte, the paternal allele is subjected to either active TET3-mediated (light blue oval) or passive (replication-dependent) DNA demethylation, resulting in an unmethylated paternal allele shortly after fertilisation (white circles). After passage through a CAST oocyte, the paternal allele ends up more highly methylated than it does after passage through in a B6 oocyte (black and white circles). **A.** CAST-oocyte derived gene product(s) (green triangle) weakly target the paternal VM-IAP allele for methylation following DNA demethylation. **B.** CAST-oocyte derived gene product(s) (maroon crescent) partially protect the paternal VM-IAP allele from active or passive DNA demethylation. **C.** CAST-oocyte derived factors contributing to the efficient demethylation of a CAST paternal genome are not optimised to demethylate a B6 paternal genome (warped light blue oval). Note: these are simplified diagrams (the extent to which VM-IAPs are fully demethylated during early development is unclear). In addition, this figure does not depict the possibility of intrauterine environment effects, which have not been ruled out.

pure B6 mice. This would allow for the differentiation of the alleles during bisulphite pyrosequencing while holding genetic background constant.

Interestingly, parent-of-origin effects were reported in a study on evolutionarily young human *Alu* elements. *Alu* elements belong to the SINE family of retrotransposons and represent the most abundant repeat class in humans, with a copy number greater than one million per genome. The study detected inter-individual variability in methylation at a set of *Alu* inserts that are polymorphic in the human population (Sandovici *et al.* 2005). Seven out of the 19 analysed *Alu* polymorphisms exhibited parent-of-origin effects, six of which showed higher methylation levels when transmitted paternally. The authors point out that all seven of these are located in centromeric or telomeric regions and propose that the parental origin-dependent methylation is related to the role that these genomic regions play in homologous chromosome pairing.

While most VM-IAPs were modulated by maternal genetic background, albeit modestly, fewer of the tested loci exhibited zygotic genetic background effects. These tended to be more striking than the maternal effects and were reminiscent of previous work on the strain-specific modulation of transgene methylation states (Allen *et al.* 1990; Engler *et al.* 1991; Sapienza *et al.* 1989; Sutherland *et al.* 2000). For VM-IAP$_{Rab6b}$, VM-IAP$_{Sema6d}$, VM-IAP$_{Fam78b}$, VM-IAP$_{Pink1}$, and VM-IAP$_{Ect2l}$, inheriting a haploid CAST genome from either parent induced hypermethylation and largely abrogated variability, suggesting that for these cases the responsible modifiers are transcribed from zygotic DNA after fertilisation rather than carried over from the CAST oocyte. Backcrossing experiments further showed that simple Mendelian inheritance of dominant alleles underlies the strain-specific modulation of a subset of solo LTR VM-IAPs, providing an experimental framework for the genetic mapping of VM-IAP modifiers.

The faithful segregation of VM-IAP$_{Rab6b}$ methylation states in the N3 generation following selection in the N2 generation confirmed that VM-IAP$_{Rab6b}$ is targeted by a single dominant CAST-derived SSM coded by the *Ssm$_{Rab6b}^{C}$* gene. VM-IAP$_{Pink1}$ and VM-IAP$_{Ect2l}$, whose sequences show more than 94% identity with the VM-IAP$_{Rab6b}$ sequence, are also SSM$_{Rab6b}^{C}$ targets. Efforts to map *Ssm$_{Rab6b}^{C}$* are in progress. N3 mice carrying *Ssm$_{Rab6b}^{C}$* will be backcrossed to B6 mice,

narrowing the location of modifier gene down to the 6.25% of the N4 genome still containing CAST DNA. N4 mice are expected to recreate the 1:1 ratio of highly to lowly methylated individuals at VM-IAP$_{Rab6b}$, VM-IAP$_{Pink1}$, and VM-IAP$_{Ect2l}$. Genomic DNA extracted from individuals of both methylation phenotypes will be run on the Giga Mouse Universal Genotyping Array (GigaMUGA) (Morgan *et al.* 2016). The GigaMUGA captures the genetic diversity found in laboratory mouse strains and probes 141,090 single nucleotide polymorphisms (SNPs) and 2,006 copy number variations. Although the assayed SNPs are overrepresented at recombination hotspots, on average they are tiled every 10.4 +/- 12.1 kb along the mouse genome. Most of the SNPs are polymorphic between B6 and CAST mice due to the significant evolutionary distance between them. The interval of CAST DNA shared by all highly methylated individuals and absent from all lowly methylated individuals is expected to contain the *Ssm$_{Rab6b}^{C}$* locus.

Apart from proving useful in the identification of VM-IAP SSMs, the polymorphic nature of VM-IAPs across mouse strains can also be exploited to gain evolutionary insight. For instance, the VM-IAP$_{Rab6b}$ sequence is present in the 129S1/SvlmJ (129Sv) genome but is not variably methylated, instead exhibiting high methylation levels similar to those observed in BC and CB hybrids (Kazachenka 2018). This indicates that 129Sv mice are likely carriers of *Ssm$_{Rab6b}^{C}$* and can be used for validation purposes following genetic mapping. In addition, since the last common ancestor between the 129Sv and CAST strains is shared with the B6 strain, *Ssm$_{Rab6b}^{C}$* must have been present in the B6 genome at some point. B6 mice may have lost *Ssm$_{Rab6b}^{C}$* via a deletion mutation or, alternatively, may possess a *Ssm$_{Rab6b}^{B}$* allele that differs enough from the *Ssm$_{Rab6b}^{C}$* allele in sequence to interfere with its ability to repress VM-IAP$_{Rab6b}$ effectively. While it is possible that CAST and 129Sv mice evolved distinct VM-IAP$_{Rab6b}$ SSMs via convergent evolution, this is unlikely due to the absence of VM-IAP$_{Rab6b}$ in the CAST genome.

The data presented in this chapter regarding SSM$_{Rab6b}^{C}$-driven modification of VM-IAP$_{Rab6b}$ are evocative of two previously published studies. The first relates to the dactylaplasia phenotype (limb malformation) in the mouse, caused by an ERV insertional mutation of the MusD subclass (*Dac^{1J}*) in the *Fbxw4* gene locus (Friedli *et al.* 2008; Kano *et al.* 2007). The phenotypic effects of the mutation are

only observed in a subset of inbred mouse strains and this is associated with the unlinked strain-specific modifier gene *Mdac* (modifier of dactylaplasia) (Johnson *et al.* 1995). In strains carrying the *Mdac* allele (including C57BL/6J), the *Dac*^{1J} MusD element is heavily methylated and the dactylaplasia phenotype is not observed. In strains lacking the *Mdac* allele, *Dac*^{1J} is unmethylated and mice exhibit dactylaplasia (Kano *et al.* 2007). The *Mdac* locus was first mapped to a 28 Mb region on chromosome 13 (Johnson *et al.* 1995) and later narrowed to a 9.4 Mb interval containing 125 candidate genes (Kano *et al.* 2007). A 2011 doctoral dissertation appears to have further narrowed the locus to a 1.7 Mb window that contains a 0.5 Mb deletion in dactylaplasia-permissive strains. The deletion spans a cluster of Krüppel-associated box (KRAB) domain zinc finger protein (KRAB-ZFP or KZFP) genes, representing likely *Mdac* candidates (Aktas 2011). This latter work has not yet been published in a peer-reviewed journal.

The second study was conducted on the HRD (heavy chain enhancer, rearrangement by deletion) transgene, originally designed to study V(D)J recombination in mice (Engler & Storb 1987). As observed for other transgenes, the methylation state of HRD is dependent on genetic background: HRD is unmethylated in DBA/2 mice and highly methylated in B6 mice (Engler *et al.* 1991). When DBA/2 mice carrying an unmethylated HRD transgene are crossed to B6 mice, the transgene becomes methylated in F1 hybrids by the single dominant B6 modifier SSM1B (strain-specific modifier 1 in B6). More than two decades after its discovery, *Ssm1b* was mapped to a 0.5 Mb interval on distal chromosome 4 and identified as *Zfp979* (Ratnam *et al.* 2014). ZFP979 belongs to a unique subfamily of KZFPs comprising an array of three zinc fingers. It is expressed in early development up to embryonic day 8.5 (E8.5) and relies on the activity of *de novo* methyltransferase DNMT3B for transgene methylation (Ratnam *et al.* 2014).

The parallels between the strain-specific targeting of *Dac*^{1J} by MDAC, of the HRD transgene by SSM1B, and of VM-IAP$_{Rab6b}$, VM-IAP$_{Pink1}$, and VM-IAP$_{Ect2l}$ by SSM$_{Rab6b}^{C}$ are suggestive of overlapping mechanisms. Neither *Mdac* nor *Ssm1b* are expected to be the *Ssm*$_{Rab6b}^{C}$ locus itself because VM-IAP$_{Rab6b}$, VM-IAP$_{Pink1}$, and VM-IAP$_{Ect2l}$ are unmethylated in B6 mice while *Dac*^{1J} and the HRD transgene are methylated in B6 mice. Nonetheless, KZFPs are likely involved in all three mechanisms. KZFPs play an established role in transposable element repression

and make up a large and rapidly evolving family of transcription factors in the mouse (reviewed in Rowe & Trono 2011), making them appealing potential VM-IAP modifiers. In the discussion chapter (Chapter 7), this hypothesis is expanded into a model that incorporates findings from multiple chapters and lays out the molecular and functional characteristics of KZFPs that render them compelling candidates as mechanistic drivers of VM-IAP metastability.

Chapter 6

Identification of a genetically-conferred epiallele

6.1 Introduction and objectives

Following the computational screen for C57BL/6J (B6) VM-IAPs, candidate loci were experimentally validated in inbred B6 mice and deemed true positives if their inter-individual methylation range was greater than 10% (Kazachenka *et al*. 2018). Chapters 2 and 5 describe these ranges in detail and showed that VM-IAP methylation levels display bounded continuous probability distributions when sampled from a B6 population (Figure 2.10 and Figure 5.2).

IAP_{Pgm1}, named after its closest coding gene *Pgm1*, was detected as a candidate VM-IAP in the screen. This IAP is a full-length element located on chromosome 5 (chr5:64,030,834-64,038,297). It contains protein coding sequences flanked by identical 5′ and 3′ LTRs of the IAPLTR2_Mm subclass (**Figure 6.1A**). The inter-individual methylation range at IAP_{Pgm1} exceeded the 10% threshold in both of the two B6 cohorts used to experimentally validate IAP_{Pgm1}. However, rather than exhibiting a continuous range of methylation levels across individuals, IAP_{Pgm1} methylation appeared to cluster into two distinct states in both cohorts(**Figure 6.1B**).

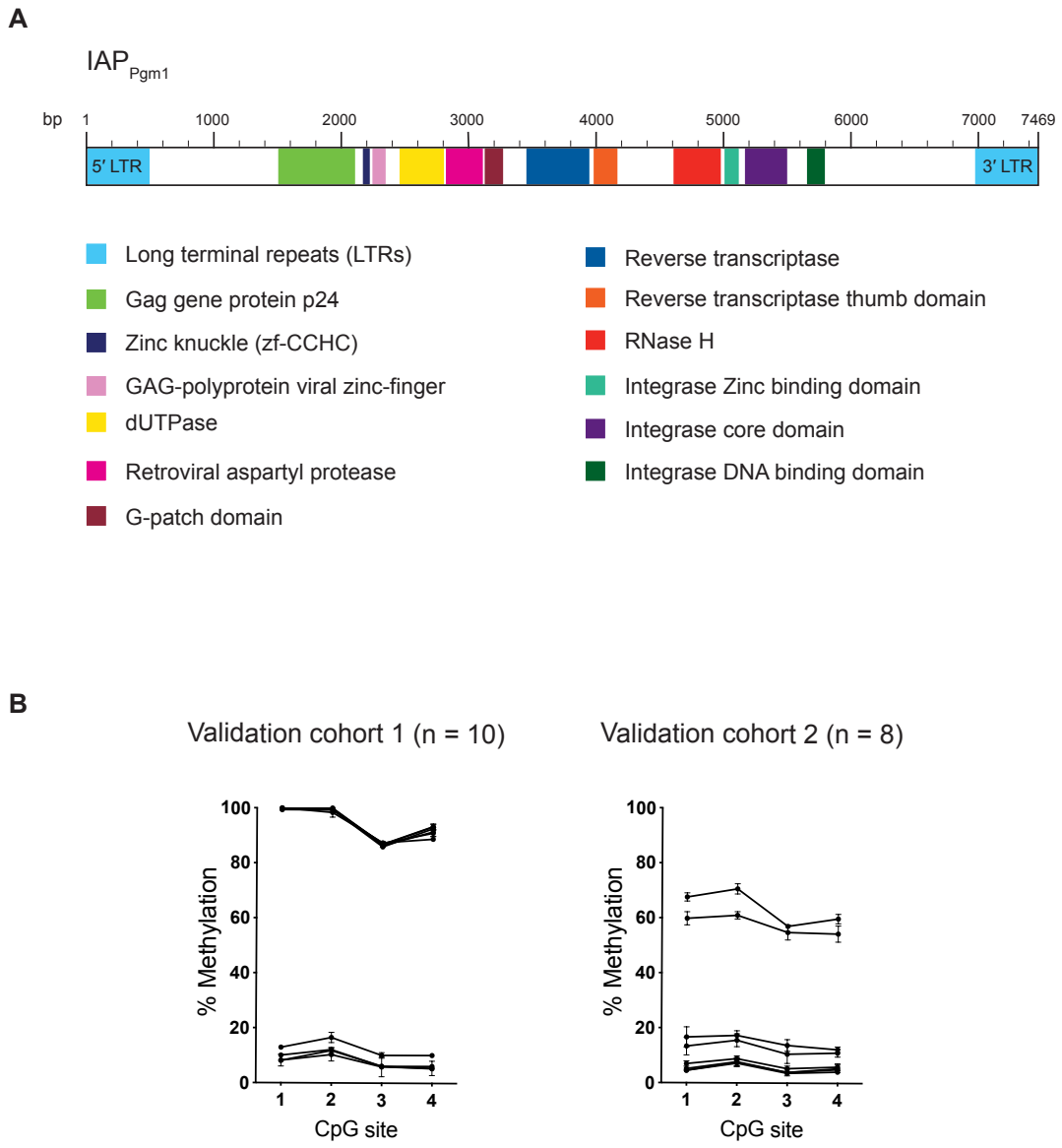


Figure 6.1: Structure and experimental validation of IAP$_{Pgm1}$. A. The genomic structure and coding potential of IAP$_{Pgm1}$, as predicted by the Pfam database (https://pfam.xfam.org). Genomic distances are drawn to scale. **B.** Methylation levels were quantified at the four most distal CpGs of the IAP$_{Pgm1}$ 5′ LTR (nearest the boundary with unique DNA). Tested samples are liver tissues collected from two sets of adult B6 mice. Both males and females are represented in each cohort. Each line represents an individual and error bars represent standard deviations between technical triplicates. These data were generated by Anastasiya Kazachenka during the experimental validation of computationally identified VM-IAP candidates (Kazachenka 2018).

This chapter aims to describe the nature, cause, and consequences of the unusual epigenetic segregation displayed by IAP_{Pgm1}, as outlined in the following objectives:

1. To characterise in more detail the methylation patterns of IAP_{Pgm1} in the B6 population;

2. To assess the heritability and developmental dynamics of IAP_{Pgm1} methylation states;

3. To determine the root cause of the epigenetic polymorphism at this locus;

4. To investigate the functional consequences of this phenomenon.

6.2 Results

6.2.1 IAP_{Pgm1} methylation is tri-modally distributed in the B6 population

To reproduce the clustering of methylation states previously observed at IAP_{Pgm1}, ear samples were collected from male and female B6 mice and assessed via bisulphite pyrosequencing. The analysis confirmed that IAP_{Pgm1} is distinct from other VM-IAPs in that its methylation pattern is tri-modal: individuals are either highly methylated (> 85% methylation), lowly methylated (< 20% methylation), or intermediately methylated (60-70% methylation) (**Figure 6.2A**). This pattern was observed in both males and females, indicating that the segregation is not sex-linked (**Figure 6.2A**).

As discussed in Chapter 1, the repetitive nature of IAPs results in primer design difficulties. To ensure that multiple IAPs are not amplified during bisulphite pyrosequencing, either the forward or the reverse primer must lie in the bordering unique region. In addition, due to the short length of PCR products amplified from bisulphite converted DNA (300-500 bp), only the distal CpGs on either end of the IAP can be interrogated. For the subset of full-length VM-IAPs whose 3′ LTR methylation levels have been assessed, no concordance in methylation was observed between the 5′ and 3′ LTRs (Kazachenka 2018). For some, only one of the LTRs is variably methylated; for others, both LTRs are variably methylated but the individual-specific methylation level is different between the two LTRs, suggesting methylation patterns are established independently at each LTR for these VM-IAPs. In contrast, the 5′ and 3′ LTR methylation levels at IAP_{Pgm1} are consistent with one another (**Figure 6.2B**). Individuals that are highly methylated at the 5′ LTR are also highly methylated at the 3′ LTR. The same consistency is observed for intermediate- and lowly methylated individuals (**Figure 6.2B**). Even though these data only pertain to the edges of IAP_{Pgm1}, the agreement in their methylation levels signals that the segregation of methylation states at IAP_{Pgm1} extends across the entire 7.5 kb element.

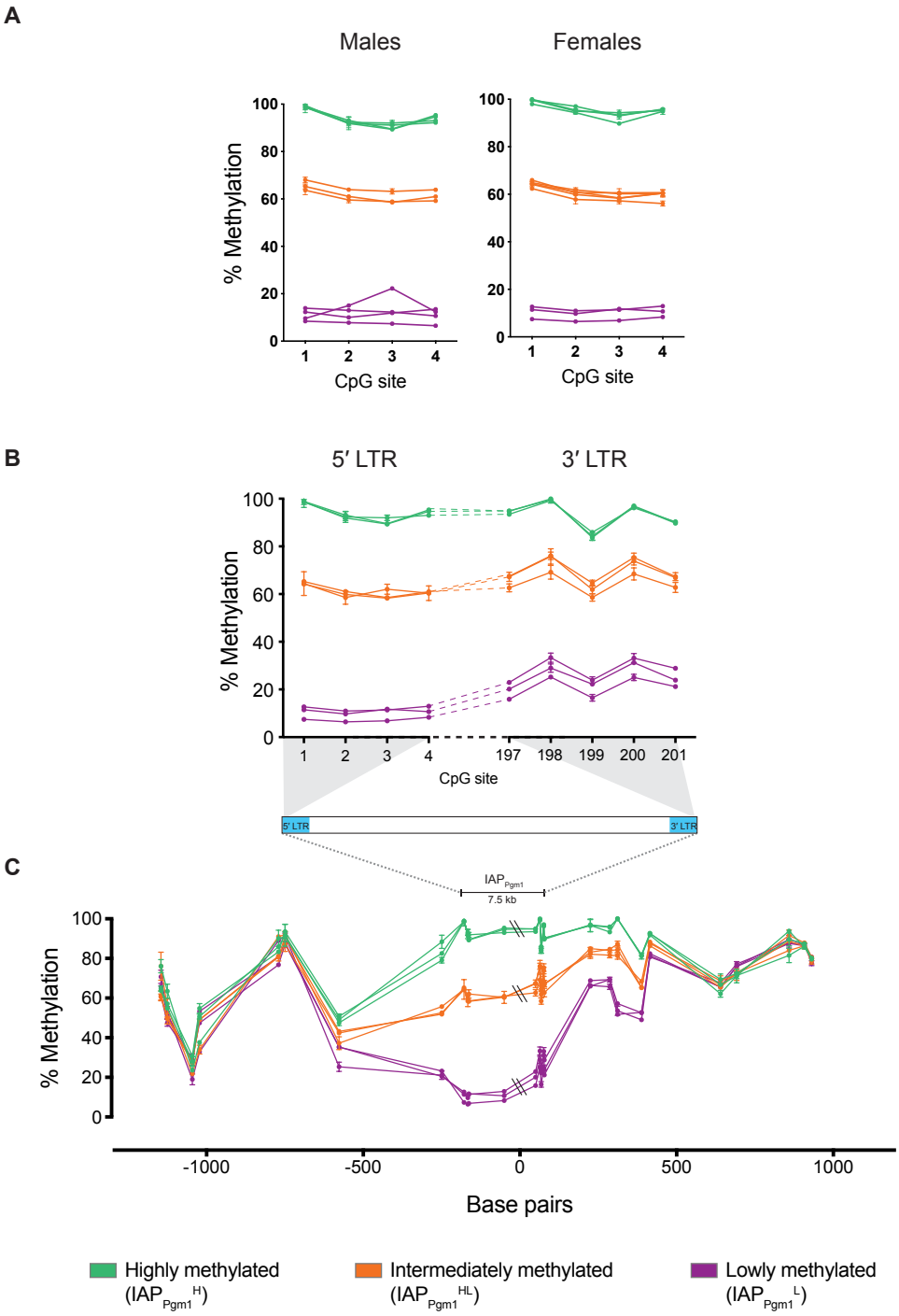


Figure 6.2: Characterisation of IAP$_{Pgm1}$ methylation states in the B6 population. A. IAP$_{Pgm1}$ methylation levels in B6 males and females segregate into three distinct states: high (green), intermediate (orange), and low (purple). **B.** IAP$_{Pgm1}$ methylation levels are consistent between the 5′ and 3′ LTRs. Dashed lines connect data from the same individual. **C.** Inter-individual methylation variation at IAP$_{Pgm1}$ collapses within 500 bp on either side of the IAP element. Methylation levels were quantified from ear samples via bisulphite pyrosequencing. Each line represents one individual and each data point represents one CpG. Error bars represent standard deviations calculated from technical triplicates.

It is unclear whether the segregation of methylation states at IAP_{Pgm1} is intrinsic to the IAP element or the result of epigenetic properties of the insertion site. To distinguish between these two possibilities, methylation was quantified at each CpG in the 1 kb window of unique DNA immediately up- and downstream of IAP_{Pgm1} (primers in

Table 8.2). As the distance from the LTR borders increased, the three distinct methylation states became less defined and ultimately collapsed approximately 500 bp out from either side of the IAP (**Figure 6.2C**). This indicates that IAP_{Pgm1} itself is likely mediating the epigenetic differences between individuals rather than having inserted into a genomic region that already manifested a clustered methylation pattern.

6.2.2 Characterization of IAP_{Pgm1} inheritance patterns

6.2.2.1 IAP_{Pgm1} methylation exhibits stable Mendelian inheritance

One of the characteristic properties of VM-IAPs is the reconstruction of metastability from one generation to the next regardless of parental methylation state (Chapter 2). To test whether this phenomenon occurs at IAP_{Pgm1}, B6 mice were epigenotyped and classified as high-, low-, or intermediately methylated. Specific parental combinations were set up for breeding and IAP_{Pgm1} methylation levels were quantified in the offspring. In stark contrast to VM-IAPs, IAP_{Pgm1} exhibited stable inheritance of methylation states. Offspring born to highly methylated parents were all highly methylated, while offspring born to lowly methylated parents were all lowly methylated (**Figure 6.3**). When one parent was highly methylated and the other was lowly methylated, all offspring were intermediately methylated (**Figure 6.3**). This suggests that high and low methylation levels are allelic states of IAP_{Pgm1}, with intermediate methylation representing co-dominant heterozygosity (if we loosely accept the use of methylation states as a phenotypic readout in this context). In line with these findings, highly, lowly, and intermediately methylated individuals at IAP_{Pgm1} were designated IAP_{Pgm1}^{HH}, IAP_{Pgm1}^{LL}, and IAP_{Pgm1}^{HL}, respectively.

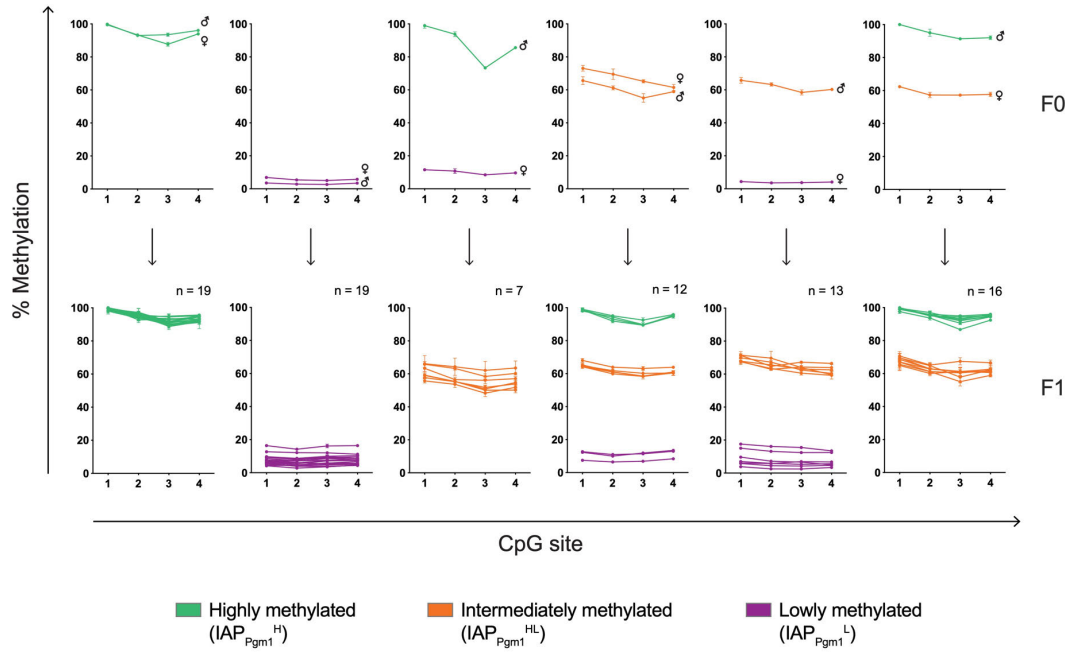


Figure 6.3: Stable Mendelian inheritance of IAP$_{Pgm1}$ methylation states. B6 mice were epigenotyped for IAP$_{Pgm1}$ methylation (F0, upper panel) and a range of parental combinations were set up for breeding. The quantification of IAP$_{Pgm1}$ methylation levels in the offspring (F1, lower panel) reveals that high (IAP$_{Pgm1}^{HH}$, green) and low (IAP$_{Pgm1}^{LL}$, purple) methylation are two allelic states of IAP$_{Pgm1}$ that are stably inherited, with intermediate methylation (IAP$_{Pgm1}^{HL}$, orange) representing heterozygosity. Methylation was quantified at the four most distal CpGs of the 5′ LTR via bisulphite pyrosequencing. Each line represents an individual. Error bars represent the standard deviation calculated from technical triplicates. Standard sex symbols are used to identify maternal and paternal methylation levels. Offspring sample sizes are shown on the upper right corner of each graph.

Additional breeding pairs were set up with IAP$_{Pgm1}^{HL}$ parents to confirm the proposed inheritance pattern. An IAP$_{Pgm1}^{HL}$ intercross produced IAP$_{Pgm1}^{HH}$, IAP$_{Pgm1}^{LL}$, and IAP$_{Pgm1}^{HL}$ offspring, as would be expected from a heterozygote intercross. Pups produced from a IAP$_{Pgm1}^{HH}$ x IAP$_{Pgm1}^{HL}$ cross were either IAP$_{Pgm1}^{HH}$ or IAP$_{Pgm1}^{HL}$, while pups produced from a IAP$_{Pgm1}^{LL}$ x IAP$_{Pgm1}^{HL}$ cross were either IAP$_{Pgm1}^{LL}$ or IAP$_{Pgm1}^{HL}$ (**Figure 6.3**). These results are consistent with the anticipated F1 outcomes of a test cross between a homozygote and a heterozygote, confirming that the segregation of IAP$_{Pgm1}$ methylation levels in the B6 population is a reflection of three epiallelic states. These results also demonstrate that the methylation state of one epiallele does not affect or exert

dominance over the methylation state of the other in a heterozygous context. The experiments described in this section suggest that IAP_{Pgm1} is inherited in a stable Mendelian fashion and should therefore be classified as a *stable*, not *metastable*, epiallele.

6.2.2.2 IAP_{Pgm1} displays an epiallele-specific maternal effect

The reciprocal hybrid breeding scheme used in Chapter 5 showed that VM-IAP methylation levels can exhibit maternal and genetic background effects. A similar strategy was adopted to determine whether IAP_{Pgm1} is also susceptible to these effects. In this case, B6 males and females were epigenotyped before mating them with CAST/EiJ (CAST) mice to ensure that IAP_{Pgm1}^{HH}, IAP_{Pgm1}^{LL}, and IAP_{Pgm1}^{HL} mice were represented. IAP_{Pgm1} is not present in the CAST genome, so BC and CB offspring carry a single maternal or paternal copy of IAP_{Pgm1} (**Figure 6.4A**). In line with previous results in pure B6 mice, hemizygous BC offspring born to a IAP_{Pgm1}^{HH} B6 dam were all highly methylated (IAP_{Pgm1}^{H}), those born to a IAP_{Pgm1}^{LL} B6 dam were all lowly methylated (IAP_{Pgm1}^{L}), and those born to a IAP_{Pgm1}^{HL} B6 mother were either IAP_{Pgm1}^{H} or IAP_{Pgm1}^{L} (**Figure 6.4B**). No intermediately methylated BC offspring were observed, further demonstrating that intermediate methylation is a product of heterozygosity.

Methylation data collected from CB individuals identified an epiallele-specific maternal effect at the IAP_{Pgm1} locus. CB offspring born to a IAP_{Pgm1}^{HH} B6 father were all highly methylated (as observed for BC offspring) but CB offspring born to a IAP_{Pgm1}^{LL} B6 father displayed an increase in methylation levels compared to lowly methylated B6 and BC individuals (**Figure 6.4C**). The increase in methylation was not consistent across CB offspring, creating a range of inter-individual methylation levels that is reminiscent of VM-IAP metastability. As explained in more detail in the discussion section of Chapter 5, the shift in methylation levels in CB IAP_{Pgm1}^{L} individuals compared to both B6 IAP_{Pgm1}^{LL} and BC IAP_{Pgm1}^{L} individuals is indicative of a strain-specific maternal effect that is likely mediated by cytoplastic factors in the CAST oocyte rather than a zygotic genetic background effect. There is an additional layer of complexity in the case of IAP_{Pgm1} because the effect is epiallele-specific: IAP_{Pgm1}^{L}

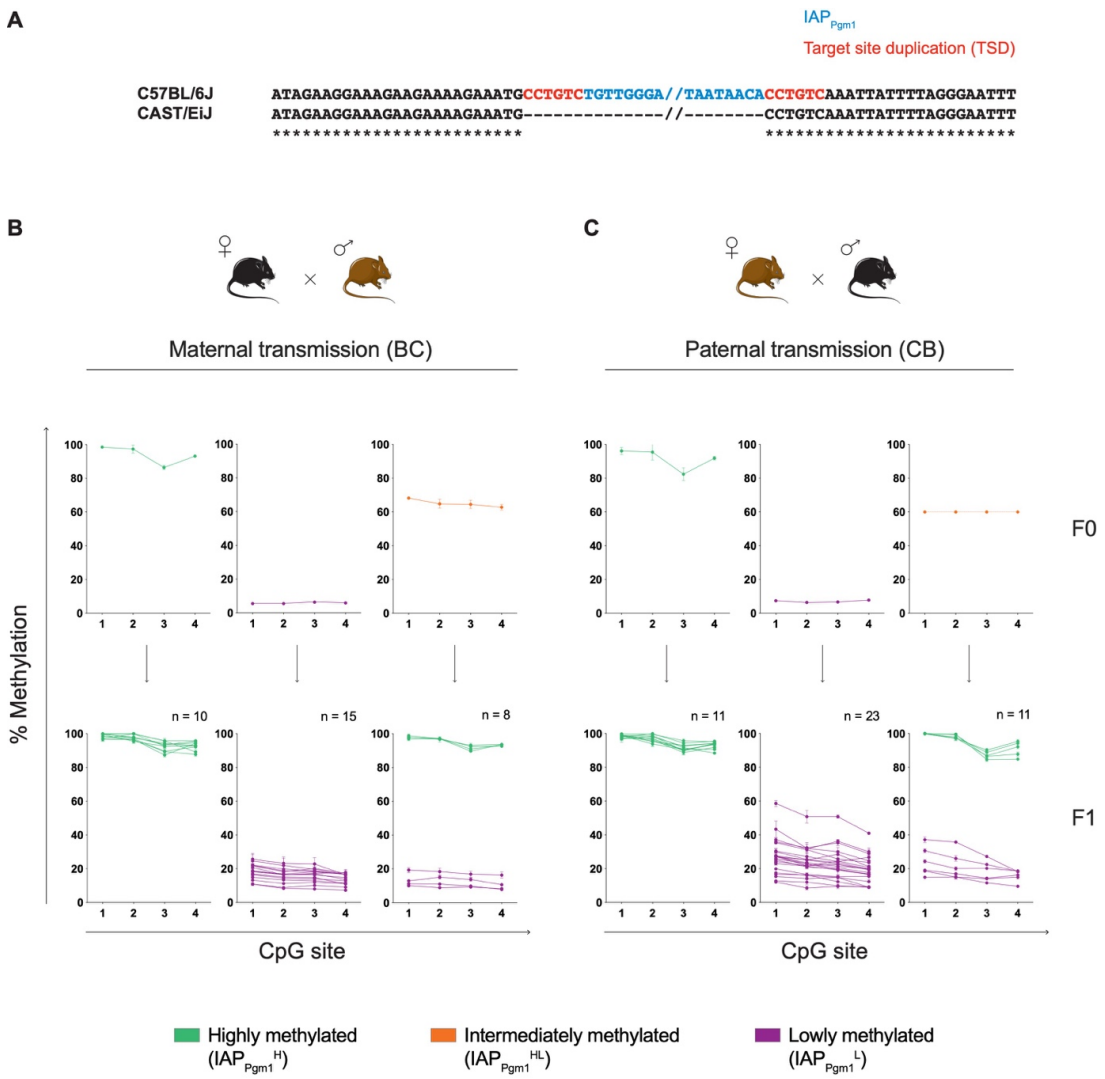


Figure 6.4: Reciprocal F1 hybrids confirm Mendelian inheritance at IAP$_{Pgm1}$ and reveal an epiallele-specific maternal effect. A. B6 and CAST sequence alignment of the IAP$_{Pgm1}$ insertion site using CLUSTAL Omega. B6 and CAST DNA sequences were extracted from the GRCm38/mm10 and CAST_EiJ_v1 assemblies, respectively, accessed through the UCSC genome browser. IAP$_{Pgm1}$ sequence is shown in blue and target site duplication sequences are shown in red. Asterisks indicate nucleotide conservation. For conciseness, the majority of the B6 sequence has been omitted from this figure; nothing is omitted from the CAST sequence. **B.** IAP$_{Pgm1}^{HH}$ (green), IAP$_{Pgm1}^{LL}$ (purple), and IAP$_{Pgm1}^{HL}$ (orange) B6 females were crossed to CAST males (F0, upper panel) and IAP$_{Pgm1}$ 5′ LTR methylation levels were quantified in hemizygous offspring (F1, lower panel). C. The same analysis was carried out in offspring born to CAST females and IAP$_{Pgm1}^{HH}$, IAP$_{Pgm1}^{LL}$, and IAP$_{Pgm1}^{HL}$ B6 males. Each line represents an individual. Error bars represent the standard deviation calculated from technical triplicates. Offspring sample sizes are shown on the upper right corner of each graph. The F0 IAP$_{Pgm1}^{HL}$ B6 male used for this experiment was produced from a cross between an IAP$_{Pgm1}^{LL}$ female and a IAP$_{Pgm1}^{HH}$ male but its methylation levels were not quantified so the dotted line represents predicted rather than measured methylation levels.

CB individuals show altered methylation levels but $IAP_{Pgm1}{}^{H}$ CB individuals do not. This may be due to the fact that the CAST oocyte appears to have an upwards-shifting effect and the baseline methylation in $IAP_{Pgm1}{}^{HH}$ individuals is already close to 100%.

The strain-specific maternal effect observed in $IAP_{Pgm1}{}^{L}$ CB individuals was interrogated in more detail by backcrossing $IAP_{Pgm1}{}^{L}$ CB males to CAST females. The additional passage of the $IAP_{Pgm1}{}^{L}$ epiallele through a CAST egg resulted in a cumulative effect, whereby N2 $IAP_{Pgm1}{}^{L}$ mice showed even higher methylation levels compared to F1 $IAP_{Pgm1}{}^{L}$ mice (**Figure 6.5**). In the next section, the methylation state of $IAP_{Pgm1}{}^{L}$ allele is investigated in the B6 male germline and shown to be hypermethylated. Therefore, since the allele is transmitted paternally in this experiment, the effect is not the direct result of methylation accumulation across generations and could instead suggest a dosage effect of a CAST modifier. It is thus possible that the CAST content of the inherited paternal genome in combination with passage through a CAST egg contributes to the cumulative effect observed in the N2 generation. Additional rounds of backcrossing are necessary to determine whether the increase in methylation will plateau at an intermediate state or whether the cumulative effect will persist transgenerationally until reaching 100% methylation.

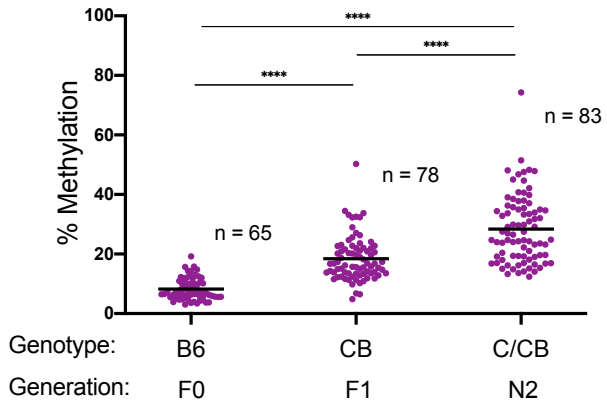


Figure 6.5: The maternal effect at $IAP_{Pgm1}{}^{L}$ is cumulative over successive generations. F1 $IAP_{Pgm1}{}^{L}$ CB males were backcrossed to CAST females to produce N2 C/CB offspring. N2 offspring were PCR genotyped for the IAP_{Pgm1} locus and only the individuals harbouring the IAP were analysed. Each data point represents the average methylation level across distal CpGs of the IAP_{Pgm1} 5′ LTR for one individual. Each sample was run in triplicate. Black lines represent group means. and Sample sizes are shown on the graph. Statistics: Kruskal-Wallis test followed by Dunn's post hoc multiple comparison test (**** $p<0.0001$).

6.2.3 Developmental dynamics of IAP_{Pgm1} methylation

6.2.3.1 IAP_{Pgm1} is differentially modified in the male and female germlines

The experiments described so far have focused on somatic IAP_{Pgm1} methylation levels, both within and across generations. The unique inheritance patterns coming out these analyses prompted the study of IAP_{Pgm1} methylation dynamics in the male and female germlines.

Adult B6 males were epigenotyped using tail DNA and six mice (two IAP_{Pgm1}^{HH}, two IAP_{Pgm1}^{LL}, and two IAP_{Pgm1}^{HL}) were selected for sperm collection. In contrast to the tri-modal segregation of methylation states observed in somatic tissues, IAP_{Pgm1} was hypermethylated in all sperm samples regardless of somatic epigenotype (**Figure 6.6A**). Thus, both the IAP_{Pgm1}^{H} and IAP_{Pgm1}^{L} alleles are targeted for repression during spermatogenesis, as is the case for VM-IAPs and most other repeat elements in the mouse genome. This suggests that the mechanisms underlying IAP_{Pgm1} methylation in the soma and the male germline are distinct. Preliminary data indicate that the DNMT3C methyltransferase is responsible for the methylation of IAP_{Pgm1} in the male germline (data not shown).

To test IAP_{Pgm1} methylation levels in the female germline, germinal vesicle (GV) oocytes were collected from the ovaries of superovulated B6 females. Oocytes collected from IAP_{Pgm1}^{HH} females were combined to form three pools of 100 oocytes each; the same was done for oocytes collected from IAP_{Pgm1}^{LL} females. IAP_{Pgm1} methylation levels were quantified for each pool and found to be reflective of somatic methylation levels: oocytes collected from IAP_{Pgm1}^{HH} females were highly methylated and oocytes collected from IAP_{Pgm1}^{LL} females were lowly methylated (**Figure 6.6B**). Therefore, a single mechanism is likely driving IAP_{Pgm1} methylation (or lack thereof) in both the soma and the female germline. Together, these experiments indicate that the maternal and paternal IAP_{Pgm1}^{L} alleles are differentially methylated in the early zygote.

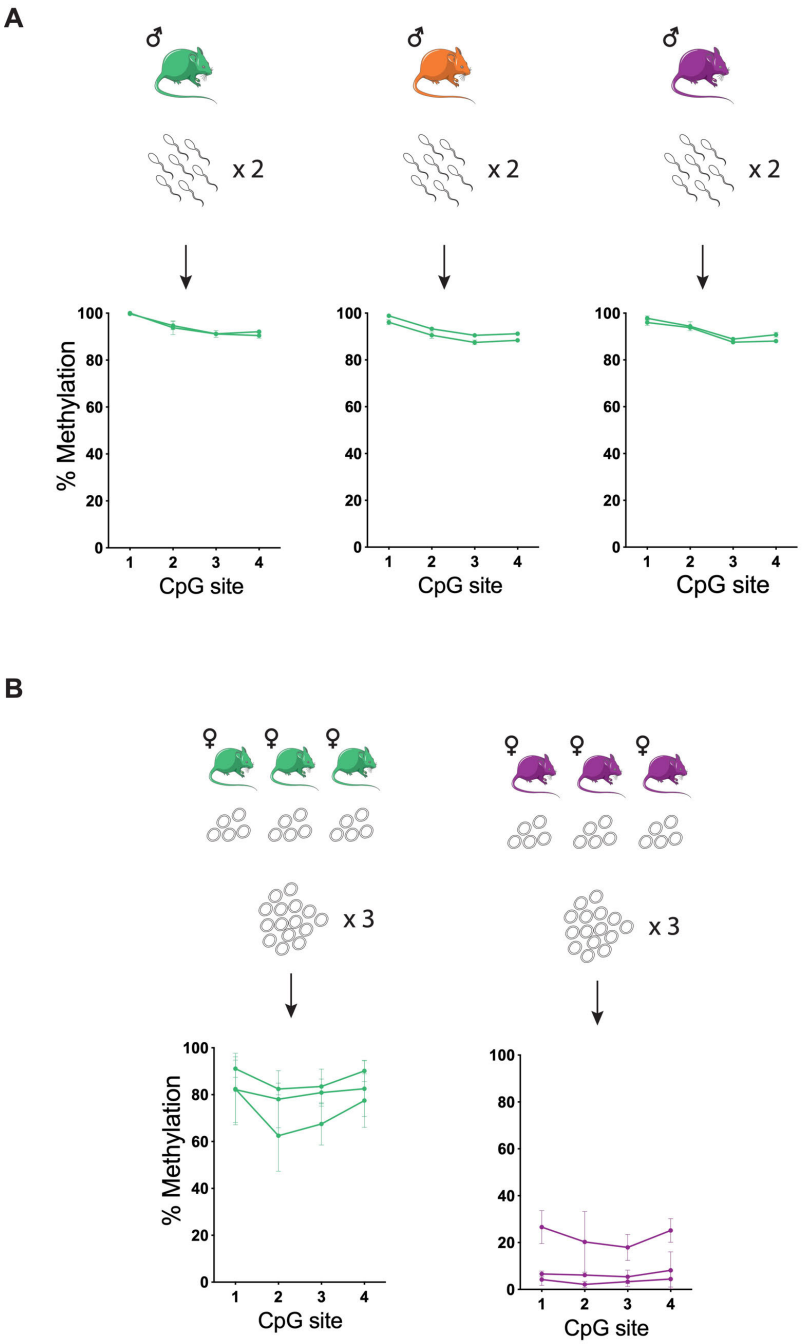


Figure 6.6: IAP$_{Pgm1}$ methylation in the germline. A. Both the IAP$_{Pgm1}^{H}$ and IAP$_{Pgm1}^{L}$ alleles are hypermethylated in the male germline. Mature sperm samples were collected and purified from the cauda epididymis of adult IAP$_{Pgm1}^{HH}$ (green, n = 2), IAP$_{Pgm1}^{LL}$ (purple, n = 2), and IAP$_{Pgm1}^{HL}$ (orange, n = 2) B6 males. The colour of the mice represents somatic methylation level while the colour of the data points and lines represents the methylation state of the sperm samples. IAP$_{Pgm1}$ 5′ LTR methylation levels were assessed separately for each sample. Each line represents data from one individual. **B.** IAP$_{Pgm1}$ methylation levels in the female germline reflect somatic methylation levels. GV oocytes were collected from superovulated IAP$_{Pgm1}^{H}$ and IAP$_{Pgm1}^{L}$ B6 females. Each line represents one pooled sample of 100 oocytes collected from multiple females. Error bars represent the standard deviation calculated from technical triplicates. Hormone injection and oocyte quality control for this experiment was conducted by Dr. Nozomi Takahashi.

6.2.3.2 *IAP$_{Pgm1}$ methylation levels are determined during preimplantation development*

To examine the behaviour of IAP$_{Pgm1}$ methylation levels after fertilisation, blastocysts were collected from the uteri of IAP$_{Pgm1}^{HH}$ and IAP$_{Pgm1}^{LL}$ B6 females bred to IAP$_{Pgm1}^{HH}$ and IAP$_{Pgm1}^{LL}$ B6 males, respectively. Littermate embryos were pooled before analysis. Blastocysts generated from IAP$_{Pgm1}^{LL}$ parents were lowly methylated at IAP$_{Pgm1}$, indicating that the paternally inherited hypermethylated IAP$_{Pgm1}^{L}$ allele is demethylated shortly after fertilisation (**Figure 6.7A**). In contrast, methylation levels in blastocysts generated from IAP$_{Pgm1}^{HH}$ parents hovered around 70% (**Figure 6.7A**). We cannot determine based on these data alone whether IAP$_{Pgm1}^{HH}$ is demethylated after fertilization and rapidly methylated again by the blastocyst stage, or whether IAP$_{Pgm1}^{HH}$ is (partially) resistant to epigenetic reprogramming during preimplantation development. Nonetheless, these results demonstrate that methylation patterns at the IAP$_{Pgm1}^{H}$ and IAP$_{Pgm1}^{L}$ epiallelic variants are specified prior to implantation.

Although the distinction in methylation states between IAP$_{Pgm1}^{HH}$ and IAP$_{Pgm1}^{LL}$ blastocysts is evident, it is noteworthy that IAP$_{Pgm1}^{HH}$ methylation levels are lower at the blastocyst stage than in adult somatic tissue. There are at least three separate explanations for the incomplete methylation in blastocysts. It is possible that the early embryos are still in the process of gaining methylation at IAP$_{Pgm1}$, given that most of the genome is at its lowest methylation level at this stage in development (Hackett & Surani 2013; Smallwood & Kelsey 2012). Alternatively, since littermate blastocysts were pooled, the intermediate methylation could be a reflection of inter-embryo variation, whereby some blastocysts have become fully methylated and others are still unmethylated. A third possibility is that the incomplete methylation is symptomatic of lower methylation levels in the developing trophectoderm (TE) that are countering higher methylation levels in the inner cell mass (ICM). To test this, ICM-derived embryonic stem cell (ESC) lines were generated from IAP$_{Pgm1}^{HH}$ and IAP$_{Pgm1}^{LL}$ blastocysts. IAP$_{Pgm1}$ methylation levels in these cell lines closely matched those observed in adult somatic tissues, with ESC lines

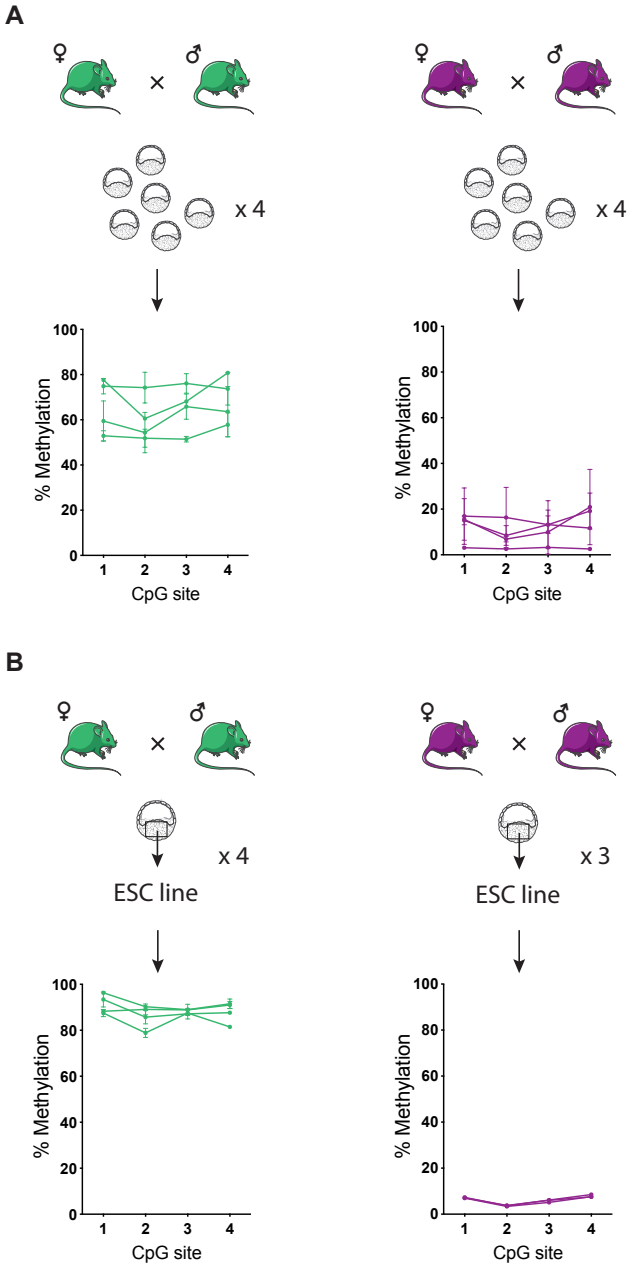


Figure 6.7: IAP$_{Pgm1}$ methylation levels are specified during preimplantation development. **A.** Pooled blastocysts generated from IAP$_{Pgm1}^{HH}$ parental crosses (green) are more highly methylated than those generated from IAP$_{Pgm1}^{LL}$ parental crosses (purple). Blastocysts were flushed from the uteri of four females of each epigenotype. Littermate blastocysts collected from the same female were pooled prior to bisulphite conversion and pyrosequencing of the IAP$_{Pgm1}$ 5′ LTR. Each line represents one blastocyst pool and error bars represent standard deviations calculated from technical triplicates. **B.** ICM-derived ESC lines generated from IAP$_{Pgm1}^{HH}$ and IAP$_{Pgm1}^{LL}$ blastocysts show high and low methylation levels at the IAP$_{Pgm1}$ 5′ LTR, respectively. Each line represents data from one cell line and each point represents the average of technical duplicates. The IAP$_{Pgm1}^{HH}$ ESC lines (green) were generated alongside Nozomi Takahashi. The IAP$_{Pgm1}^{LL}$ ESC lines (purple) were generated by Dr. Takahashi for a previous study, so the epigenotype of the B6 male and female used to produce the blastocysts from which the cells were derived is not known (light purple).

derived from IAP$_{Pgm1}^{HH}$ blastocysts exhibiting close to 100% methylation (**Figure 6.7B**). While ESCs do not necessarily reflect *in vivo* ICM biology, this result is suggestive of differential methylation between the ICM and the TE at the IAP$_{Pgm1}$ locus.

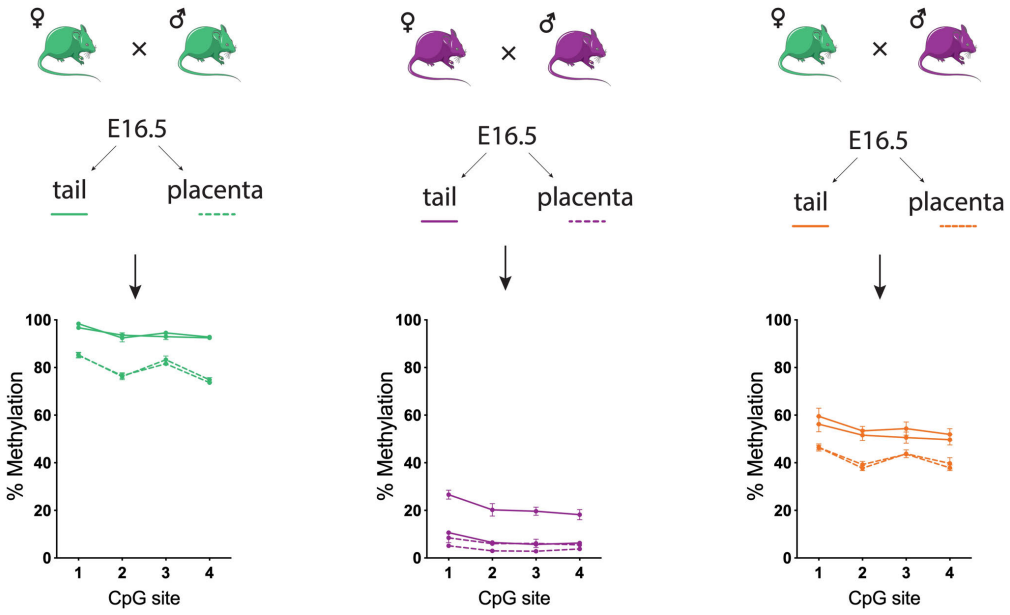


Figure 6.8: Placental IAP$_{Pgm1}$ methylation levels are hypomethylated but retain the tri-modal distribution. IAP$_{Pgm1}$ 5′ LTR methylation levels were quantified in matched E16.5 tail (solid lines) and placenta (dashed lines) samples collected from IAP$_{Pgm1}^{HH}$ (green, n = 2), IAP$_{Pgm1}^{LL}$ (purple, n = 2), and IAP$_{Pgm1}^{HL}$ (orange, n = 2) embryos. Each line represents one individual and error bars represent standard deviations calculated from technical triplicates.

To explore this possibility further, IAP$_{Pgm1}$ methylation levels were compared between E16.5 tail and placenta samples dissected from IAP$_{Pgm1}^{HH}$, IAP$_{Pgm1}^{LL}$, and IAP$_{Pgm1}^{HL}$ embryos. Methylation levels were lower in placenta than in tail for all epigenotypes, consistent with the results from section 3.X in Chapter 3 and with previous studies on global placental hypomethylation (**Figure 6.8**; Ehrlich *et al.* 1982; Schroeder *et al.* 2015). While this finding is compatible with the ESC and blastocyst data, IAP$_{Pgm1}^{HH}$ blastocyst methylation levels were lower than IAP$_{Pgm1}^{HH}$ E16.5 placenta methylation levels, suggesting that a combination of the explanations outlined above are likely at play. Of note, even though placental IAP$_{Pgm1}$ methylation levels are hypomethylated compared to their embryonic

counterparts, they retain the tri-modal segregation of epigenotypes. This observation supports the hypothesis that IAP_{Pgm1} methylation states are established prior to the first cell lineage specification event segregating the TE and the ICM. Furthermore, the faithful distinction in IAP_{Pgm1} methylation levels between IAP_{Pgm1}^{HH} and IAP_{Pgm1}^{LL} ESC lines demonstrates that ESCs are a viable *in vitro* model to study the mechanisms underlying the acquisition of methylation at this locus.

6.2.4 The methylation state of IAP_{Pgm1} is genetically determined

The stable transmission of IAP_{Pgm1} methylation states from one generation to the next is more reminiscent of classical genetics than of epigenetic inheritance. It is likely that the IAP_{Pgm1}^{H} and IAP_{Pgm1}^{L} epialleles are conferred by a spontaneous genetic mutation in the B6 population, either in the IAP element itself or in a gene involved in its epigenetic regulation. As the latter possibility would likely require whole genome sequencing of multiple mice, the former possibility was investigated in the first instance.

To search for local single nucleotide polymorphisms (SNPs) between IAP_{Pgm1}^{HH} and IAP_{Pgm1}^{LL} individuals, primers were designed to amplify the entirety of IAP_{Pgm1} via PCR with the intent to sequence the resulting PCR products via Sanger sequencing. As mentioned above, the large number of IAPs in the B6 genome with high sequence identity necessitates that at least one PCR primer is designed to target the bordering unique region of the element. Two sets of primers were designed to amplify each half of the IAP_{Pgm1} element with an 183 bp overlap between the two fragments (**Figure 6.9A**, pink P1 and yellow P2 arrows; **Table 8.1**). Genomic DNA from three IAP_{Pgm1}^{HH} and three IAP_{Pgm1}^{LL} mice was used as template for the PCR. Unexpectedly, while the IAP_{Pgm1}^{HH} reactions amplified both of the anticipated PCR products, the IAP_{Pgm1}^{LL} reactions amplified neither (**Figure 6.9B**). This was the first line of evidence that pointed to a genetic difference between IAP_{Pgm1}^{HH} and IAP_{Pgm1}^{LL} individuals. In addition, since neither the P1 nor the P2 reaction yielded a band when using IAP_{Pgm1}^{LL} samples as template, the genetic polymorphism could not be due to a single SNP.

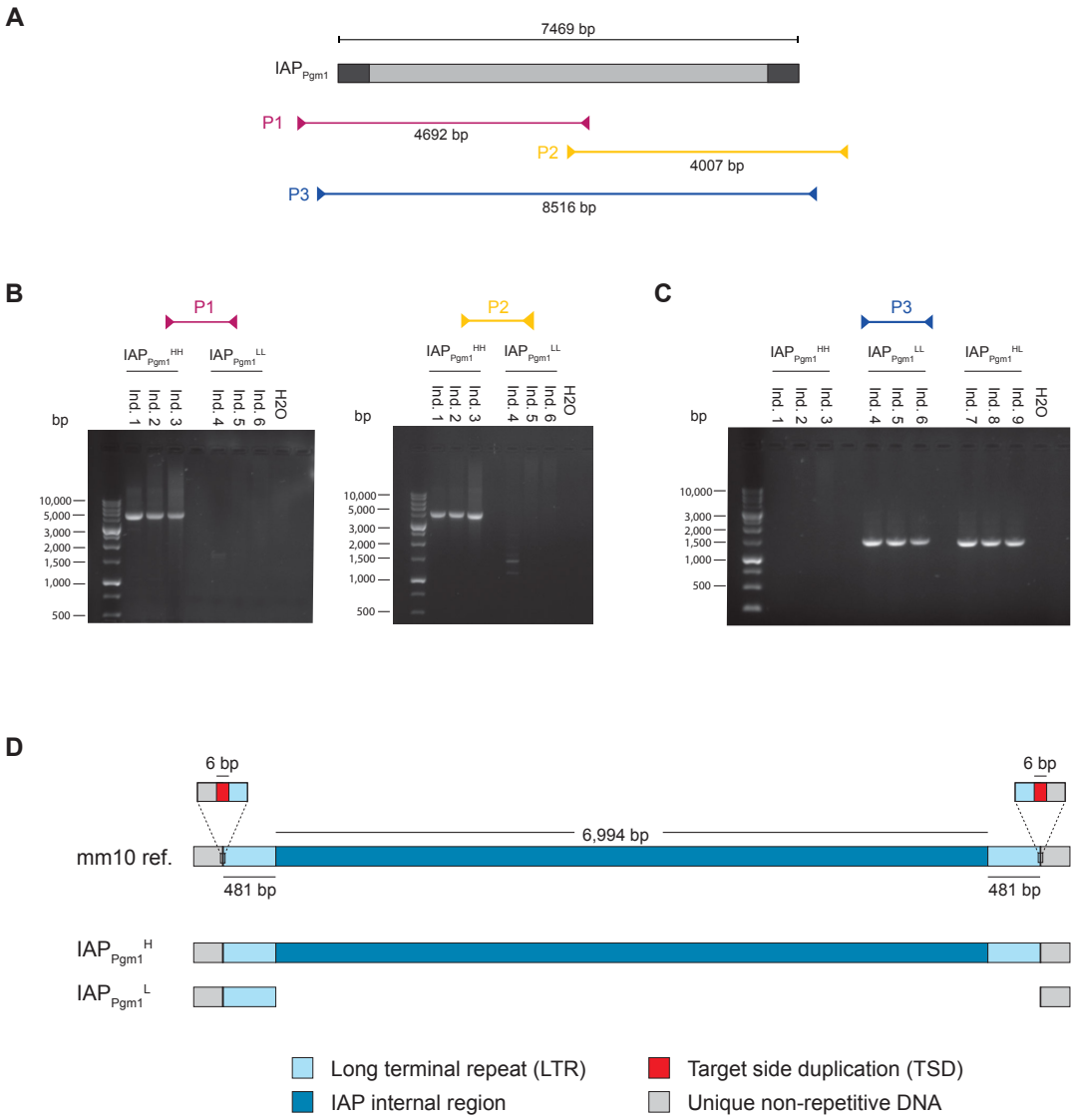


Figure 6.9: The IAP$_{Pgm1^L}$ allele is the product of an inter-LTR recombination event. A. PCR primer map. **B.** Agarose gel electrophoresis of PCR products amplified from IAP$_{Pgm1^{HH}}$ and IAP$_{Pgm1^{LL}}$ DNA using P1 and P2 primer pairs. **C.** Agarose gel electrophoresis of PCR products amplified from IAP$_{Pgm1^{HH}}$, IAP$_{Pgm1^{LL}}$, and IAP$_{Pgm1^{HL}}$ DNA using the P3 primer pair. **D.** Schematic representation of the alignment of the IAP$_{Pgm1}$ mm10 reference sequence and the assembled IAP$_{Pgm1^H}$ and IAP$_{Pgm1^L}$ sequences following Sanger sequencing. The LTRs (light blue), internal region (dark blue), target site duplication (TSD; red), and unique non-repetitive DNA (grey) are shown to scale. The alignment of the IAP$_{Pgm1^L}$ solo LTR to the 5′ LTR of the IAP$_{Pgm1^H}$ allele is arbitrary and could equivalently be shown aligned to the 3′ LTR. The base-resolution alignment can be found in Appendix A, **Figure A.8**.

To determine whether a more substantial genetic difference is present in IAP_{Pgm1}^{LL} individuals, primers were designed to target the unique bordering regions on either side of the IAP element (**Figure 6.9A**, blue P3 arrows; **Table 8.1**). Based on the GRC38/mm10 genome assembly, these primers should amplify an 8,516 bp fragment. No bands were observed following the use of this primer pair on IAP_{Pgm1}^{HH} DNA, likely because the PCR settings were not optimal for amplification of such a large DNA fragment (**Figure 6.9C**). In contrast, both IAP_{Pgm1}^{LL} and IAP_{Pgm1}^{HL} DNA templates gave rise to bands of approximately 1,500 bp in length, indicating the presence of a large 7kb deletion between the two P3 primer annealing sites on the IAP_{Pgm1}^{L} allele (**Figure 6.9C**).

Following gel purification, the two larger fragments amplified from IAP_{Pgm1}^{HH} DNA using the P1 and P2 primer pairs and the smaller fragment amplified from IAP_{Pgm1}^{LL} DNA using the P3 primer pair were used as templates for Sanger sequencing. For the larger fragments, sequencing primers were interspersed at regular intervals across the IAP element (**Table 8.4**). For the smaller fragment, the same primers used for PCR were used for sequencing. The resulting assembled sequences for the IAP_{Pgm1}^{H} and IAP_{Pgm1}^{L} alleles were aligned using CLUSTAL Omega. The alignment revealed that the IAP_{Pgm1}^{L} allele contains a single IAP LTR while the IAP_{Pgm1}^{H} allele perfectly matches the full-length GRC38/mm10 reference sequence (**Figure 6.9D**; Appendix A, Figure A.8). The IAP_{Pgm1}^{L} solo LTR and the IAP_{Pgm1}^{H} full-length IAP are both flanked by the same target site duplications (TSDs) and the IAP_{Pgm1}^{L} solo LTR is identical in sequence to both the 5′ and 3′ LTRs of the full-length IAP_{Pgm1}^{H} (**Figure 6.9D**). Thus, this mutation was almost certainly generated by a recent inter-LTR recombination event in the inbred B6 population (**Figure 6.10**).

Inter-LTR recombination events occur when the identical or near-identical LTR sequences engage in ectopic homologous recombination, resulting in the formation of solo LTRs (reviewed in Jern & Coffin 2008). Since both IAP_{Pgm1} allelic variants were detected in B6 samples sent to us from collaborators in mainland Europe and North America (data not shown), the recombination event probably occurred at The Jackson Laboratory (JAX) or at another C57BL/6J-distributing facility. This type of proviral excision event is common: more than half of the IAPs in the mouse genome are solo LTRs (Shimosuga *et al.* 2017). Less common is the unmethylated state of the identified solo LTR, considering that

the vast majority of the ~5,000 solo LTRs in the B6 genome are highly methylated (Shimosuga *et al.* 2017). This raises questions regarding the mechanisms that underlie the epigenetic silencing of IAP_{Pgm1} as well as the functional consequences of a solo LTR left unmodified.

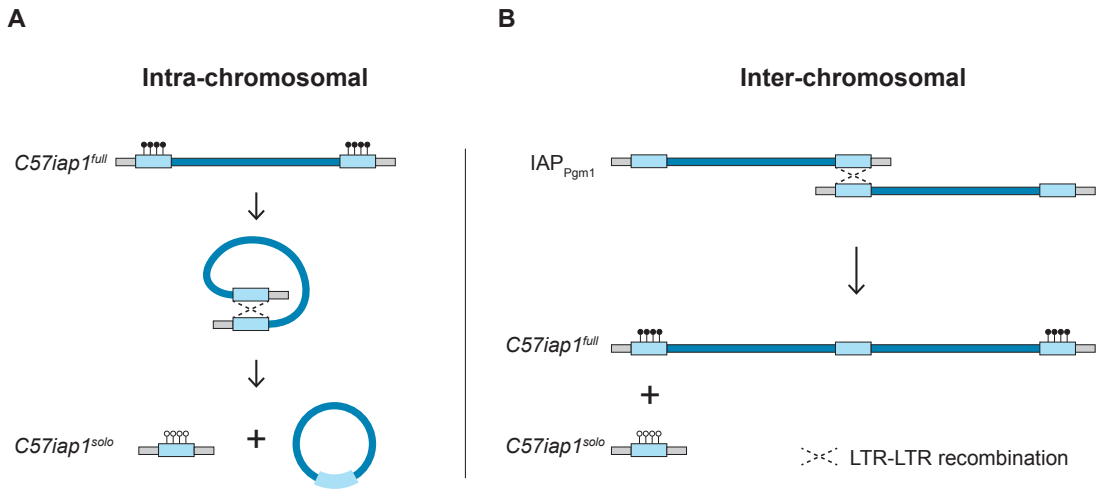


Figure 6.10: Possible mechanisms of LTR-LTR recombination leading to a solo LTR at the *C57iap1* locus. A. Intra-chromosomal homologous recombination between the 5′ and 3′ LTRs of the repeat element, resulting in the formation of a solo LTR and a circularised unintegrated viral fragment. **B.** Inter-chromosomal recombination between identical LTRs on sister chromatids or homologous chromosomes, producing a solo LTR on one DNA strand and a tandem duplication on the other. Diagrams adapted from Seperack *et al.* 2006.

As a caveat to this experiment, it should be mentioned that the lack of amplification of an 8,516 bp fragment using the P3 primer pair on IAP_{Pgm1}^{HH} DNA leaves open the possibility that the full-length IAP element on the IAP_{Pgm1}^{H} allele is duplicated in tandem. As shown in **Figure 6.10**, LTR-LTR recombination could occur intra-chromosomally between the 5′ and 3′ LTRs of the same repeat element, or inter-chromosomally between LTRs on sister chromatids or homologous chromosomes. Given the large number of solo LTRs in the genome and the relatively small proportion of reported tandem IAP duplications, the former mechanism seems to occur more frequently. However, additional experiments are required to determine whether the absence of amplification is due to unfavourable PCR conditions or evidence of a tandem duplication on the IAP_{Pgm1}^{H} allele.

The sequencing experiments described in this section demonstrate that the IAP_{Pgm1}^{L} epiallele identified in this chapter was caused by an inter-LTR recombination event. Thus, the allelic variants IAP_{Pgm1}^{H} and IAP_{Pgm1}^{L} will henceforth be referred to as *C57BL/6J IAP 1^{full}* (*C57iap1^{full}*) and *C57BL/6J IAP 1^{solo}* (*C57iap1^{solo}*), respectively. This nomenclature is intended to facilitate the naming of other polymorphic repeat elements in the future, both in B6 and in other inbred mouse strains. A PCR genotyping assay was developed to easily distinguish *C57iap1^{full}* from *C57iap1^{solo}* individuals in subsequent experiments (primer sequences in **Table 8.1**).

6.2.5 Functional consequences of the *C57iap1* polymorphism

6.2.5.1 C57iap1 allelic variants are associated with adjacent gene expression in a tissue-specific manner

IAP insertions can influence neighbouring gene expression in a number of ways. Intragenic IAPs can lead to aberrant transcript formation by providing alternative splice sites or causing premature transcript termination. Intergenic IAPs can induce the formation of chimeric transcripts initiated at a cryptic promoter in the IAP LTR and may also act as enhancers (reviewed in Gagnier *et al.* 2019). These effects are often dependent on the epigenetic properties of the IAP or the chromatin dynamics of the insertion site. For example, the *A^{vy}* and *Axin^{Fu}* IAPs only drive ectopic *Agouti* or *Axin* expression when they are unmethylated. It is therefore possible that *C57iap1* allelic variant influences the expression of genes in its genomic vicinity (*Pgm1*, *Rell1*, *Tbc1d1*, and *5830416l19Rik*). The position of these genes in relation to *C57iap1* is depicted in **Figure 6.11A**.

As the original name of *C57iap1* implies, *Pgm1* (*Phosphoglucomutase-1*) is the closest coding gene to *C57iap1*, lying 55 kb downstream of the 3′ LTR. PGM1 catalyses the interconversion between glucose 1-phosphate and glucose 6-phosphate, but its role is secondary to that of the predominant PGM isozyme, PGM2. *Pgm1* is lowly expressed in the placenta and intestine while *Pgm2* is highly expressed in a more diverse set of tissues, including heart, liver, mammary gland,

and placenta. Confusingly, the nomenclature of *Pgm1* and *Pgm2* is inversed in the mouse compared to other vertebrate species (the human orthologue of mouse *Pgm1* is *PGM2*, and vice versa). PGM1 deficiency in humans is associated with glycogen storage disease and a congenital disorder of glycosylation. Clinical symptoms include growth retardation, cardiomyopathy, and bifid uvula at birth, among others (Beamer 2015). *PGM2* has been associated with metabolic disease in GWAS studies (Timmons *et al.* 2018).

With respect to other genes, *Tbc1d1* (*TBC1 domain family, member 1*) is located 120 kb downstream of *C57iap1* and is expressed in testis and kidney. TBC1D1 is a Rab-GTPase-activating protein with roles in insulin signalling and obesity suppression (Dokas *et al.* 2016; Roach *et al.* 2007). The *Pgm1* and *Tbc1d1* genes are therefore both involved in metabolic processes. *Rell1* (*RELT-like protein 1*) codes for a membrane protein and lies 61 kb upstream of *C57iap1*. It is expressed in the placenta and adrenal gland. *5830416l19Rik* (*Mus musculus RIKEN cDNA 5830416l19 gene*) is a long non-coding RNA expressed at low levels in the thymus and is of particular interest because it lies just 7 kb downstream of *C57iap1*. Little is known about the function of *Rell1* and *5830416l19Rik*. The information provided in this paragraph was extracted from the National Center for Biotechnology Information (NCBI) gene database (Geer *et al.* 2009).

The effect of the *C57iap1* polymorphism on the expression of these genes was assessed via quantitative real-time PCR (qRT-PCR). Expression was quantified relative to *Hprt* in liver, brain (cortex), thymus, and placental tissues collected from *C57iap1*^{full} and *C57iap1*^{solo} individuals, genotyped via PCR prior to dissection. Liver, brain, and thymic tissues were dissected from the same set of adult B6 males. Whole placentas were dissected at embryonic day E16.5. *Pgm1*, *Tbc1d1*, and *Rell1* were expressed in all tissues examined. *5830416l19Rik* transcripts were only detected in the thymus (**Figure 6.11B**).

Pgm1 and *Rell1* expression levels in the thymus and placenta were significantly greater in *C57iap1*^{solo} individuals compared to *C57iap1*^{full} individuals, which suggests that the unmethylated solo LTR is associated with increased expression

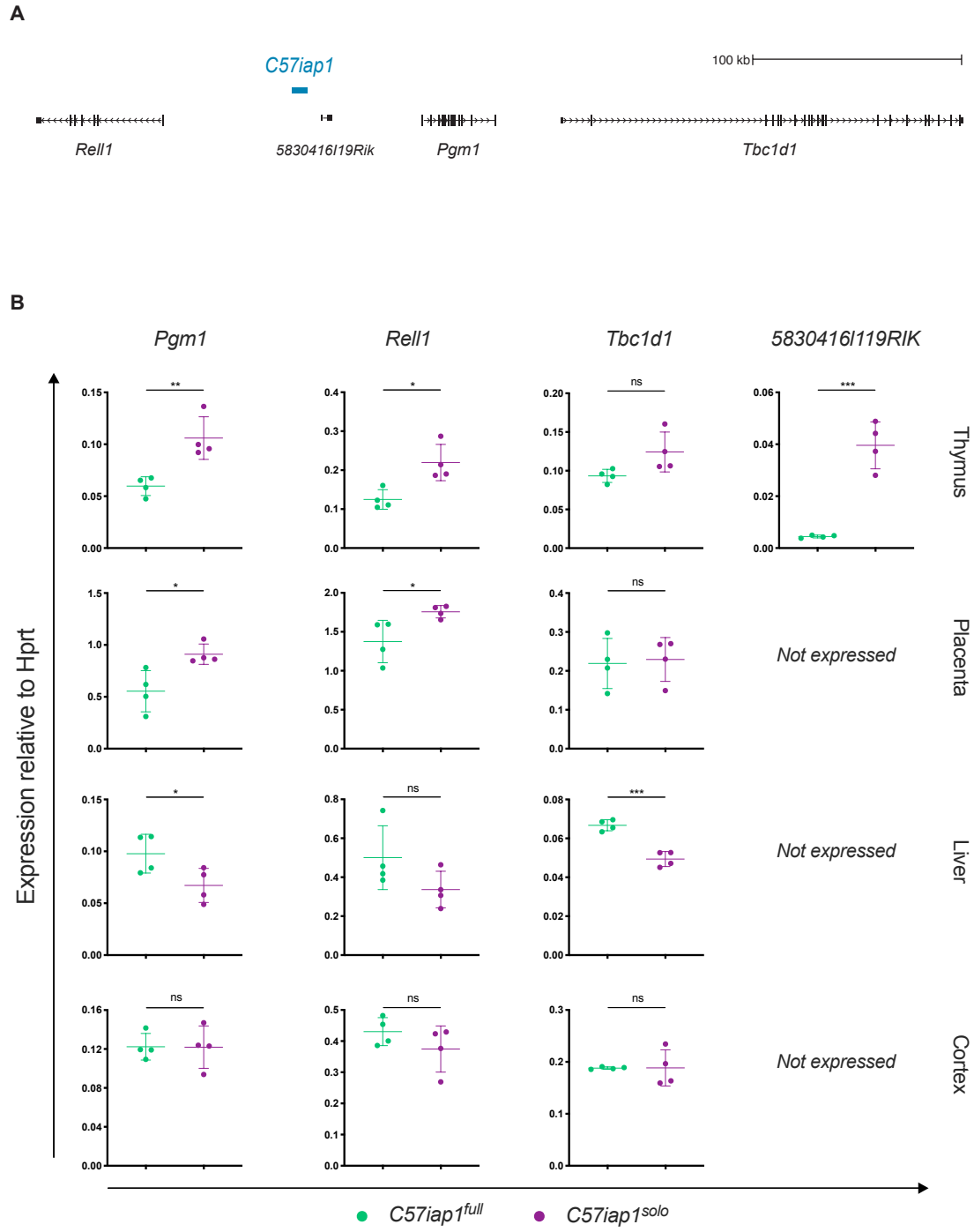


Figure 6.11: *C57iap1* allelic variant influences neighbouring gene expression in a tissue-specific manner. A. Gene map surrounding the *C57iap1* locus. Gene transcripts (black) were extracted from the University of California, Santa Cruz (UCSC) Genome Browser (Haeussler *et al.* 2019). *C57iap1* is shown in purple. Diagram drawn to scale. **B.** qRT-PCR expression data of *Pgm1*, *Rell1*, *Tbc1d1*, and *5830416l19Rik* in *C57iap1*^{full} (green) and *C57iap1*^{solo} (purple) thymus, placenta, liver, and brain tissues (n = 4 per tissue per genotype). *5830416l19Rik* is only expressed in the thymus. Relative expression was normalised to *Hprt1* expression and calculated using the ΔCt method. Statistics: unpaired t tests (*p < 0.05; ** p < 0.005; *** p < 0.0005; ns, not significant).

(**Figure 2.8B**). This inverse relationship between IAP methylation and adjacent gene expression is a similar to that observed at metastable epialleles (Kazachenka *et al.* 2018; Morgan *et al.* 1999; Rakyan *et al.* 2003). *Tbc1d1*, the furthest in distance from *C57iap1*, did not display significant differences in expression in these tissues, suggesting the *C57iap1* polymorphism influences closer host gene expression in *cis* (**Figure 2.8B**). *5830416l19Rik* expression levels, barely detected in *C57iap1*^{full} thymus samples, were significantly higher in *C57iap1*^{solo} thymus samples (**Figure 2.8B**). The liver expression data were at odds with the results from other tissues: *Pgm1* and *Tbc1d1* expression levels were significantly lower in *C57iap1*^{solo} individuals (**Figure 2.8B**). The mechanism by which this occurs is unclear. No significant differences in expression were observed for any of the genes in the cortex (**Figure 2.8B**). Together, these data show that the *C57iap1* polymorphism is associated with altered adjacent gene expression in a tissue-specific manner.

6.3 Discussion

This chapter describes the identification and characterisation of *C57iap1* as a polymorphic repeat element in the ostensibly isogenic B6 mouse strain caused by an inter-LTR recombination event. Findings demonstrate that *C57iap1* allelic variants display distinct IAP methylation profiles which are associated with differences in adjacent gene expression. The phenotypic implications of this mutation remain undefined. The tissue-specific nature of the qRT-PCR results warrants the analysis of additional tissue types to better understand the expression dynamics associated with each allele. It should be noted that most of the expression differences reported here do not exceed a fold change, consistent with the lack of conspicuous phenotypes distinguishing *C57iap1*^{full} and *C57iap1*^{solo} mice, at least in the absence of environmental stress. Given that both *Pgm1* and *Tbc1d1* are involved in metabolic pathways, it is possible that a phenotypic difference would emerge between *C57iap1*^{full} and *C57iap1*^{solo} mice in response to metabolism-altering exposures such as a high fat diet. Preliminary findings indicate that plasma metabolite levels are comparable between *C57iap1*^{full} and *C57iap1*^{solo} mice in an unaltered environment (data not shown).

It is unclear whether the *C57iap1*-mediated effects on gene expression are due to the presence of the 7 kb deletion itself or a consequence of the unmethylated state of the *C57iap1*^{solo} allele. It would be useful to compare *C57iap1* to recombination-induced polymorphisms that do not result in hypomethylation of the solo LTR, but the identification of such loci with both variants still present in the B6 population is difficult. Historically, inter-LTR recombination events have been identified due to phenotypic reversions of ERV-induced mutations, whereby the phenotypic effect of the insertion is reversed following proviral excision and solo LTR formation (Bultman *et al.* 1994; Seperack *et al.* 2006; Stoye *et al.* 1988). However, methylation of the produced solo LTRs was never interrogated in these studies.

The *C57iap1* polymorphism is a valuable endogenous system to study the mechanisms underlying IAP repression. Unlike the vast majority of solo LTRs in the B6 genome, *C57iap1*^{solo} is unmethylated. This indicates that the *C57iap1* LTR itself is not targeted for repression, perhaps due to a binding site mutation. It follows that the methylation of *C57iap1*^{full} in the soma relies on the recognition of its internal

sequence. In addition to their contrasting DNA methylation profiles, the *C57iap1*^{full} and *C57iap1*^{solo} alleles likely exhibit different histone modification landscapes. An important next step will be to conduct ChIP-qPCR experiments to quantify histone modification occupancy at each allele. Marks of interest include H3K9me3 (heterochromatic and recruited to repeat elements by KRAB-ZFPs), H3K27ac (euchromatic and used as a marker of enhancers), and H4K20me3 (most prominent histone mark found at IAP LTRs) (Martens *et al.* 2005; Rowe & Trono 2011).

The breeding experiments outlined in this chapter showed that passage through a CAST egg promotes partial methylation of the *C57iap1* LTR. The possible mechanisms by which CAST oocyte cytoplasmic factor(s) might contribute to increased methylation of paternally inherited IAPs were discussed in Chapter 5 and are equally relevant in the context of the *C57iap1*^{solo} allele. It is noted that intrauterine environment effects have not been ruled out in this case either. Since there is notable inter-individual methylation variation at the *C57iap1*^{solo} allele in CB hybrids, *C57iap1*^{solo} is a *bona fide* VM-IAP in the context of the CB genetic background. The *C57iap1* locus is therefore not only of value in investigating the consequences of repeat element structure on epigenetic repression, but also in the study of methylation acquisition at metastable epialleles. The cumulative gain in methylation upon consecutive round of backcrossing to CAST females is once again reminiscent of transgene behaviours reported on decades ago (Allen *et al.* 1990; Kearns *et al.* 2000).

Using a method called TEPBAT (Target Enrichment after Post-Bisulphite Adaptor Tagging), a 2017 study analysed the DNA methylation levels of more than 8,500 IAP LTR elements in the B6 genome in both tail and sperm samples (Shimosuga *et al.* 2017). In line with other studies, the vast majority of IAP LTRs were highly methylated, with 93% of them showing more than 80% methylation in tail samples. Of the 14 elements that showed less than 20% methylation in tail, 10 of these were solo LTRs (mostly of the IAPLTR2_mm subtype, the same as *C57iap1*). The *C57iap1* LTR exhibited 21% methylation in this analysis and was classified as a hypomethylated 3′ LTR of a full-length IAP. The identification of the *C57iap1* polymorphism suggests Shimosuga and colleagues actually detected the hypomethylated *C57iap1*^{solo} allele. It is conceivable that some, if not all, of the other hypomethylated LTRs of full-length IAPs reported in this study are in fact

solo LTRs formed from recent inter-LTR recombination events. In a comparative analysis between the tail and sperm datasets, the authors found that the hypomethylated IAPs in tail and sperm were non-overlapping and enriched in different IAP subfamilies. Most of the hypomethylated IAPs in tail were of the IAPLTR2_Mm subclass while most of those in sperm were of the IAPEy_LTR subclass. This is in line with the observed hypermethylation of *C57iap1*^{solo} in sperm and highlights the divergence of mechanisms that drive transposable element silencing in the soma and in the male germline.

The results in this chapter demonstrate that cryptic genetic diversity in inbred mouse population can have functional repercussions with implications for experimental outcomes. The identification of the *C57iap1* polymorphism should serve as a cautionary tale for researchers working with inbred mice, particularly in the field of epigenetics. Had the accuracy of the mm10 genome assembly been blindly accepted or had the sequencing experiments not extended beyond the portion of the LTR probed for methylation, the conclusions of this work would have been misinterpreted. We are reminded that our current reference genomes harbour large gaps and inaccuracies, largely due to mapping difficulties associated with the repeat genome and heterochromatic centromeric regions. It is therefore currently impossible to completely rule out genetic effects contributing to the outcomes of inbred mouse experiments. Inter-LTR recombination events are likely an underappreciated source of genetic variation, as most studies attribute the lack of uniquely mapped reads in internal portions of repeat elements to complications that arise from the non-uniqueness of their sequence, and do not consider the possibility that the sequence has been lost from a particular repeat locus. The advent of long-read sequencing and improved computational pipelines will vastly improve these annotations. A recent study in humans developed a pipeline to capture dimorphic human ERVs (HERVs) resulting from inter-LTR recombination events and detected dozens of previously unidentified candidates (Thomas *et al.* 2018). As the cost of high throughput sequencing decreases, it is predicted that whole-genome resequencing of experimental animals or test subjects will become a routine practice in the process of eliminating genetic confounders.

Chapter 7

Discussion

7.1 Summary of findings

In the past decade, both scientists and society at large have been captivated by the possibility that environmentally induced phenotypes can be transmitted across generations via epigenetic mechanisms. The A^{vy} mouse model is often cited as the best described instance of mammalian epigenetic inheritance and as a sensitive biosensor of environmental compromise. The assumption has been that other regions in the mouse genome with similar genetic and epigenetic properties likely behave in the same way. This thesis characterises a recently identified set of variably methylated endogenous retroviruses of the IAP subclass (VM-IAPs) and in doing so reveals that this premise is far too simplistic. VM-IAP$_{Gm13849}$ stood out as the only VM-IAP to exhibit memory of maternal methylation level in a manner analogous to the A^{vy} locus. Even then, the effect size was small, raising questions about its biological significance. The overall lack of heritability observed at VM-IAPs demonstrates that variable methylation at repeat elements is not necessarily associated with epigenetic inheritance and warns against extrapolating from isolated examples. Chapter 2 draws attention to the faithful re-establishment of VM-IAP metastability from one generation to the next regardless of parental methylation level. This was a universal feature of all tested VM-IAPs, suggesting epigenetic reprogramming must occur to some degree at these loci.

The developmental dynamics of VM-IAP methylation states were explored in Chapter 3. Male germline hypermethylation was identified as another universal property of metastable epialleles. This included the A^{vy} locus, at which the methylation level in sperm had previously been reported to reflect somatic methylation state (Blewitt *et al.* 2006; Rakyan *et al.* 2003). This finding points to a mechanistic divergence between the soma and the germline, perhaps indicative of a selective disadvantage associated with incomplete retrotransposon silencing in DNA that is passed on to the next generation. In the same chapter it was shown that the recently discovered DNA methyltransferase DNMT3C is required for VM-IAP methylation in the male germline and preliminary evidence was presented suggesting that this enzyme may also regulate somatic VM-IAP methylation levels.

Methylation analyses on placental samples, ESCs, and blastocysts in Chapter 3 indicated that VM-IAP metastability is most likely acquired prior to implantation and mitotically maintained from some point thereafter. In Chapter 5 it was demonstrated that this process is dependent on both maternal and zygotic genetic background. This result has implications for the mechanism driving methylation stochasticity at VM-IAPs (discussed in detail in the following sections).

The data presented in Chapter 4 suggest that VM-IAPs are selectively susceptible to environmental perturbation. Abnormal folate metabolism shifted VM-IAP methylation levels and was associated with changes in expression of VM-IAP-neighbouring genes. In contrast, no significant effects were detected following perinatal exposure to an obesogenic diet or to the endocrine disruptor BPA. A longitudinal ageing analysis was conducted on VM-IAPs to assess the effect of ageing on metastable epialleles. The results indicated that VM-IAP methylation levels are stable across the murine lifespan, with only small increases in methylation detected for a subset of loci. The findings from this chapter question the use of metastable epialleles as all-encompassing biosensors of environmental stress but leave open the possibility that they may play a role in mediating phenotypic outcomes under certain environmental conditions.

Chapter 6 described the identification and characterisation of *C57iap1* as a novel polymorphic IAP in the C57BL/6J (B6) genome. This element is annotated as a full-length IAP in the GRCm38/mm10 reference genome, complete with proviral DNA flanked by identical LTR sequences. A recent inter-LTR recombination event led to the formation of a solo LTR, which created two allelic variants in the inbred B6 population: $C57iap1^{full}$ and $C57iap1^{solo}$. As observed for the vast majority of IAPs, $C57iap1^{full}$ is highly methylated in the soma. In contrast, $C57iap1^{solo}$ is unmethylated in somatic tissues, which sets it apart from other solo LTRs in the genome. The genetic and/or epigenetic differences between $C57iap1^{full}$ and $C57iap1^{solo}$ lead to differential expression of adjacent genes. A developmental analysis of *C57iap1* methylation revealed that both the $C57iap1^{full}$ and $C57iap1^{solo}$ alleles are hypermethylated in the male germline but their respective high and low methylation patterns are present in both oocytes and preimplantation embryos. Of note, $C57iap1^{solo}$ gains methylation in a variable and cumulative manner after passage through a CAST oocyte. This enables the use of this allele as a model to study the acquisition of metastability at repeat elements.

7.2 Conceptualising a KZFP-driven model for VM-IAP metastability

In Chapter 5 it was demonstrated that a CAST-specific modifier, $SSM_{Rab6b}{}^{C}$, promotes hypermethylation of $VM\text{-}IAP_{Rab6b}$, $VM\text{-}IAP_{Pink1}$, and $VM\text{-}IAP_{Ect2l}$. This finding is reminiscent of previous studies on the Dac^{1J} retroelement and the HRD transgene (Kano *et al.* 2007; Ratnam *et al.* 2014). In both of those cases, strain-specific methylation patterns were attributed to a single dominant strain-specific modifier (SSM). The Dac^{1J} modifier *Mdac* has yet to be identified but was mapped to an interval on chromosome 13 containing a cluster of Krüppel-associated box (KRAB) domain zinc finger proteins, also known as KZFPs (Aktas 2011; Johnson *et al.* 1995; Kano *et al.* 2007). The HRD modifier gene *Ssm1b* was identified as *Zfp979* (Ratnam *et al.* 2014). The known role of KZFPs in targeting repressive epigenetic states to transposable elements (TEs) and the resemblance between these studies and the data presented here on $Ssm_{Rab6b}{}^{C}$ renders KZFPs prime VM-IAP modifier candidates. This section describes features of the KZFP family that

support this hypothesis, drawing on findings from this dissertation to devise a model that places KZFPs at the centre of the mechanism underlying the establishment of stochastic methylation at VM-IAPs.

7.2.1 Pertinent KZFP characteristics

KZFPs play an important role in TE repression (Ecco *et al.* 2017). Their zinc finger arrays recognise and bind DNA motifs and their KRAB domain recruits KAP1, which in turn recruits a range of repressing factors that induce H3K9 trimethylation and DNA methylation (Rowe & Trono 2011). KAP1 binding is enriched at IAP 5′ UTRs (downstream of the 5′ LTR) and KAP1 deletion causes IAP upregulation (Rowe *et al.* 2010). Specific murine KZFPs have been found to regulate ERVs, including ZFP809, ZFP819, and the paralogues ZFP932 and Gm15446 (Ecco *et al.* 2016; Tan *et al.* 2013; Wolf & Goff 2009). Thus, KZFP/KAP1-mediated TE repression is an established mechanism by which methylation machinery could be recruited to VM-IAPs.

The results in Chapter 5 showed that VM-IAPs can be targeted by more than one modifier and that these modifiers can be strain-specific. This rules out epigenetic regulators that are highly conserved across mouse strains as potential VM-IAP SSMs. A recent study placed the number of murine KZFPs in the hundreds, making them the largest and most diverse transcription factor family in the mouse genome (Kauzlaric *et al.* 2017). As proven by the *Ssm1b* case, the rapid evolution and expansion of KZFPs allow for polymorphisms between mouse strains, especially between distantly related strains such as B6 and CAST. The specificity of KZFP binding to a target site relies on four amino acids within each zinc finger, so strain-specific point mutations in these key residues could have important implications for the binding kinetics of a given KZFP (Elrod-Erickson *et al.* 1998).

KZFPs exhibit a distinctive clustered organisation within the genome. A majority of human KZFPs are located in six clusters on chromosome 19 (Lukic *et al.* 2014), and mouse KZFPs have been assigned to 18 clusters distributed across 11 chromosomes (Kauzlaric *et al.* 2017). This unique spatial patterning has been attributed to segmental duplications within clusters, resulting in high sequence

similarity among KZFPs in the same cluster. Murine KZFP clusters are depleted of other gene types and enriched for transposable elements. New cluster sites are thought to be seeded via retrotransposition of KZFP genes and amplified by positive selection of the inserted KZFP pseudogene (Kauzlaric *et al.* 2017). A binding site analysis of mouse paralogues ZFP932 and Gm15446 showed that they both target ERVK elements but are each enriched at different ERVK subfamilies (Ecco *et al.* 2016). Data in Chapter 5 showed that some VM-IAPs, such as VM-IAP$_{Sema6d}$ and VM-IAP$_{Fam78b}$, are likely targeted by more than one SSM. It is feasible that these VM-IAPs are simultaneously bound by a KZFP and its recently formed paralogue. It is also possible that VM-IAPs seemingly targeted by a single SSM, such as VM-IAP$_{Rab6b}$, VM-IAP$_{Pink1}$, and VM-IAP$_{Ect2l}$, are actually bound by multiple KZFPs that lie in a single cluster and mimic a single-gene inheritance pattern due to high genetic linkage.

If SSM$_{Rab6b}^{C}$ is a KZFP, the duplicative nature and high TE density of KZFP clusters may complicate its identification. The mapping experiment would likely identify a KZFP cluster rather than a single gene. Unfortunately, the sequence of KZFP clusters is largely incomplete in the current CAST reference genome. Ongoing efforts by our lab aim to fill these gaps and others (representing approximately 20% of the CAST genome) using nanopore long-read sequencing technology, which will be of great value for this experiment. Depending on the nature of the mapped interval, additional backcrossing may have to be conducted to generate more meiotic recombination events and narrow down the modifier locus, followed by expression vector experiments in B6 ES cells on gene candidates. By way of illustration, the ZFP cluster on distal chromosome 4 that harbours *Ssm1b* (or *Zfp979*) is plagued by large gaps in the reference genomes of all strains apart from B6. Therefore, while the modifier has been successfully identified in the B6 genome, the nature of the equivalent locus in the DBA/2 genome is yet to be resolved (Ratnam *et al.* 2014). As a result, Ratnam and colleagues have not yet determined whether *Ssm1b* is absent from mouse strains that do not methylate the HRD transgene or whether allelic variants of *Ssm1b* are present in these strains but differ in sequence such that they are unable to target the transgene for repression (e.g. via the acquisition of mutation(s) in DNA-binding residues).

The findings from Chapters 2 and 3 combined with previous work on A^{vy}, $Axin^{Fu}$, and VM-IAPs suggest that individual-specific methylation states at metastable epialleles are acquired during preimplantation development. This requires VM-IAP modifiers to be expressed during this developmental stage. A large number of human and mouse KZFPs are highly expressed in embryonic stem cells (ESCs), suggesting that KZFPs are active in the early embryo (Corsinotti *et al.* 2013; Imbeault *et al.* 2017; Kauzlaric *et al.* 2017). In addition, findings from Chapter 5 demonstrated that VM-IAPs are subject to maternal effects that are potentially driven by oocyte cytoplasmic factors, and at least three KZFPs (ZFP57, ZFP708, and ZFP36L2) have been characterised as maternal effect genes expressed in the oocyte (Li *et al.* 2008; Ramos *et al.* 2004; Seah *et al.* 2019).

A survey of *Ssm1b* (or *Zfp979*) expression in the germline found that *Ssm1b* is expressed in both male and female germ cells (Ratnam *et al.* 2017). More recently, it has become clear that a subset of KZFPs are expressed in differentiated cell types and adult tissues (Imbeault *et al.* 2017; Kauzlaric *et al.* 2017). Jessica Elmer in our team has recently identified tissue-specific VM-IAPs previously considered false positives from the VM-IAP screen. It is possible that these are regulated by KZFPs that are expressed in a cell type-specific manner. Overall, KZFP expression profiles are consistent with the expected expression patterns of VM-IAP modifiers.

The KZFP features outlined above were mostly discussed in the context of VM-IAP SSMs. The model described below puts forth the idea that KZFPs are central not only to the strain-specific effects observed at VM-IAPs, but also to the more general mechanism underlying methylation metastability. Although the model was conceived based on the results presented herein, it is noted that this thesis does not contain experiments on KZFPs themselves. As such, the model should be regarded as a framework for hypothesis testing in the future rather than as a summary of findings.

7.2.2 The LAB-KZFP model

The low-affinity binding of KZFPs (LAB-KZFP) model assumes that the acquisition of VM-IAP methylation states occurs in two phases: (1) a short establishment phase in early development during which VM-IAPs become methylated in a stochastic and locus-specific manner, and (2) a longer maintenance phase during which the methylation levels established in the first phase are inherited mitotically over successive cellular divisions, resulting in roughly equal methylation levels across tissues of a single individual. The mechanism by which methylation patterns are inherited upon DNA replication is well-characterised and involves the recognition of hemimethylated DNA at replication forks by the E3 ubiquitin-protein ligase UHRF1 followed by recruitment of the maintenance DNA methyltransferase DNMT1 (reviewed in Greenberg & Bourc'his 2019). The initial establishment phase of VM-IAP methylation states is more puzzling from a mechanistic perspective. The description below focuses on this phase.

As the name implies, the proposed model revolves around low-affinity binding of KZFPs. The model hypothesises that stochastic methylation arises when VM-IAP sequences are weakly recognised by KZFPs during preimplantation development. When a given KZFP binds to a target IAP with high affinity, the IAP is expected to become methylated in all cells, which is what occurs at most IAPs in the genome. A spontaneous mutation in the sequence coding for the KZFP DNA-binding residues or in the binding site of the target IAP sequence could decrease the strength of the binding interaction without completely abrogating it. The probability P of this binding event occurring (and subsequently leading to VM-IAP methylation) would be directly related to the frequency distributions of VM-IAP methylation levels in the B6 population (**Figure 7.1**). In fact, because sampling distributions become narrower as sample size increases, the width of the frequency distribution of each VM-IAP may reflect the number of cells in the growing embryo subject to P. This could potentially be used to estimate the time point in development when methylation is established at each locus.

The locus-specificity of VM-IAP methylation levels is in line with this model: the interaction between a KZFP and a VM-IAP would occur independently at

each locus, even in cases where the same KZFP targets more than one VM-IAP. In strains that do not harbour the mutation, the target IAP would be faithfully repressed, as reported for some VM-IAPs in Chapter 5. Although many factors have the potential to influence the binding kinetics at each VM-IAP (outlined below), the model predicts that the sequence of an IAP and that of its KZFP modifier are the most important drivers of methylation metastability at VM-IAPs.

A number of factors are expected to influence the probability *P* that a VM-IAP becomes methylated (**Figure 7.1**). Interactions between some or all of these are likely collectively contributing to VM-IAP methylation states. The main factors considered here are:

1. *Number of KZFP modifiers.* If more than one low-affinity KZFP is able to bind a given VM-IAP, *P* would incorporate the binding probabilities of all relevant KZFPs.

2. *Expression patterns.* The timing and level of gene expression varies across KZFPs. Higher expression levels for a more prolonged period of time would increase *P* and result in higher methylation levels at some VM-IAPs than at others, as the data suggest. KZFP expression patterns may also dictate the time point in early development when VM-IAPs are targeted. For instance, experiments on VM-IAP$_{Fam78b}$ in Chapter 2 showed that this VM-IAP is unmethylated at the blastocyst stage but eventually reaches relatively high methylation levels, perhaps because the KZFP modifiers of VM-IAP$_{Fam78b}$ are expressed later in development but exhibit higher affinity or greater expression levels than other VM-IAP-modifying KZFPs. The finding from Chapter 5 showing that VM-IAP$_{Fam78b}$ is not targeted by the same SSMs as VM-IAP$_{Rab6b}$ and VM-IAP$_{Sema6d}$ is in agreement with this. Expression patterns of maternal-effect KZFP modifiers may also underlie instances of epigenetic inheritance across generations observed at the A^{vy}, $Axin^{Fu}$, and VM-IAP$_{Gm13849}$ loci. For example, a KZFP modifier of A^{vy} may be expressed at higher levels in oocytes of pseudoagouti females than in oocytes of yellow females, leading to a bias towards maternal coat colour in the offspring.

3. *VM-IAP structure and sequence composition.* The structure of VM-IAPs is diverse, ranging from solo LTRs to truncated or full-length elements. KZFPs are able to bind a number of DNA motifs along the full length of ERVs, so the structure of an IAP would directly affect the number of KZFPs that recognise it. In addition, the transcription factors recruited by the various transcription factor binding sites located within VM-IAPs could interfere with KZFP/VM-IAP binding kinetics. For instance, VM-IAPs are enriched for CTCF binding (Kazachenka *et al.* 2018). CTCF is methylation-sensitive and could potentially act as an antagonist to a methylation-promoting KZFP modifier.

4. *Genomic context.* VM-IAP insertions are dispersed across the genome in both intronic and intergenic regions and are surrounded by a range of chromatin landscapes. KZFP-mediated repression involves the direct or indirect recruitment of heterochromatin factors by KAP1, including the histone deacetylating NuRD/HDAC complex, the H3 K9 methyltransferase SETDB1, heterochromatin protein 1 (HP1), and, eventually, DNA methyltransferases. These factors would be more likely to interact with KAP1 in heterochromatic portions of the genome, whose condensed state would also limit the binding accessibility of other VM-IAP-binding proteins such as CTCF. The heterochromatin-spreading HUSH complex might also be involved in this process, contributing to position effect variegation (PEV) at VM-IAPs. Thus, VM-IAPs located close to heterochromatin may have higher average methylation levels than those located in open chromatin.

5. *Environmental perturbation.* Any environmental change that influences the activity of KZFP-KAP1 interactors could result in a shift in VM-IAP methylation levels. For example, the A^{vy} and $Axin^{Fu}$ alleles are susceptible to dietary methyl supplementation (see section 1.4.5) and VM-IAP methylation levels are altered in mice with abnormal folate metabolism (Chapter 4). These environmental interventions have the potential to alter the availability of methyl groups used by SETDB1 and DNMTs to repress KZFP targets. It is possible that the repercussions of this are more profound for regions that are bound by low-affinity KZFPs.

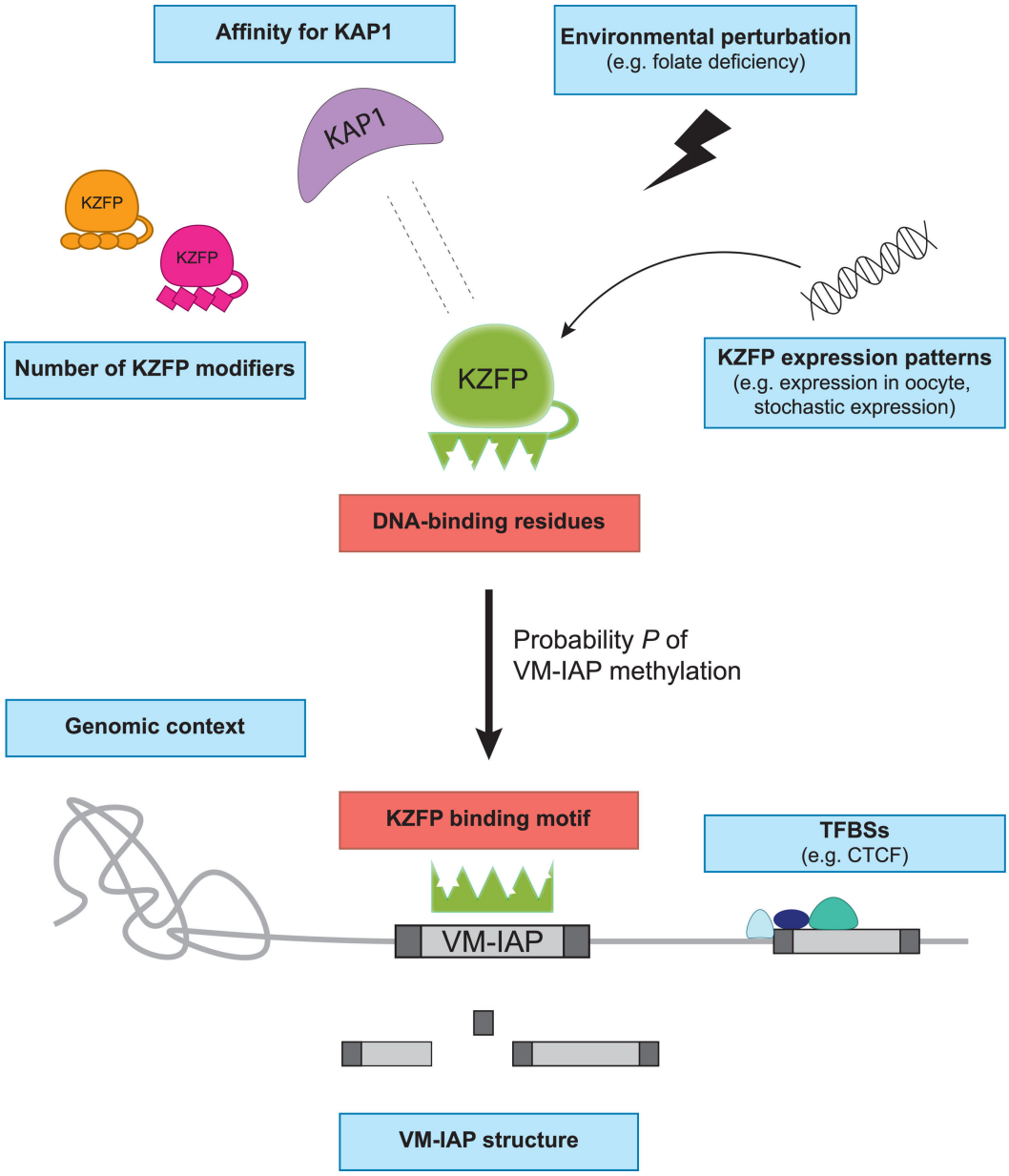


Figure 7.1: The LAB-KZFP model. Based on the model, the probability *P* that a VM-IAP becomes methylated relies primarily on the KZFP modifier DNA-binding residues and/or on the KZFP binding motif in the VM-IAP (red boxes). The nicks in the binding region represent low affinity. Secondary factors expected to influence *P* are labelled in blue boxes. Abbreviations: KAP1, KRAB-associated protein 1; KZFP, KRAB zinc finger protein; TFBS, transcription factor binding site.

The VM-IAP$_{Tfpi}$ and VM-IAP$_{Mbnl1}$ pair represent an interesting case study to evaluate the LAB-KZFP model. The sequences of these two elements are identical, which supports the idea that DNA sequence is an important factor in the acquisition of variable methylation. Based on the model, the exact same low-affinity KZFP(s) act on the two VM-IAPs. The binding dynamics would occur independently at each locus, in line with previous observations that an individual can be highly methylated at VM-IAP$_{Tfpi}$ and lowly methylated at VM-IAP$_{Mbnl1}$, or vice versa (Kazachenka *et al.* 2018). The population frequency distributions of both VM-IAP$_{Tfpi}$ and VM-IAP$_{Mbnl1}$ are right-skewed (**Figure 2.10**), but the distribution at VM-IAP$_{Tfpi}$ extends to lower methylation levels compared to that of VM-IAP$_{Mbnl1}$. This difference could be due to the genomic context of their insertion sites; perhaps VM-IAP$_{Mbnl1}$ is in a more heterochromatic region than VM-IAP$_{Tfpi}$. This would support the idea that genomic context can influence VM-IAP methylation states but plays a secondary role to VM-IAP sequence.

Similarly, in Chapter 5 it was shown that VM-IAP$_{Rab6b}$, VM-IAP$_{Pink1}$, and VM-IAP$_{Ect2l}$ share more than 94% sequence identity and are targeted for repression by the same CAST-derived modifier SSM$_{Rab6b^C}$, again highlighting that sequence is central to methylation metastability. The four other VM-IAPs that exhibit high sequence identity with VM-IAP$_{Rab6b}$, as yet untested, are also expected to show *Ssm*$_{Rab6b^C}$-dependent methylation patterns. If any of them do not, there will be an opportunity for a sequence comparison analysis aimed at identifying relevant binding site(s) within the solo LTRs. The LAB-KZFP model would predict that *Ssm*$_{Rab6b^C}$ is a high-affinity KZFP allele present in the CAST genome and that a lower affinity allele of the same gene, *Ssm*$_{Rab6b^B}$, is present in the B6 genome and drives stochastic methylation in that strain. However, it is also possible that a low-affinity KZFP that is unrelated to SSM$_{Rab6b^C}$ is binding to these VM-IAPs in B6 mice. As noted for VM-IAP$_{Tfpi}$ and VM-IAP$_{Mbnl1}$, the chromatin landscapes of the VM-IAP$_{Rab6b}$, VM-IAP$_{Pink1}$, and VM-IAP$_{Ect2l}$ insertion sites likely play a role in defining the maxima and minima of their methylation ranges across individuals.

Results presented in Chapter 3 indicate that VM-IAP methylation levels are variable across ESC lines. Subsequent work carried out by Amir Hay in our team has shown that cell populations derived from a single ESC clone do not maintain the methylation state of the cell they came from (unpublished). In addition, the

methylation range of each VM-IAP across ESC populations appears to mirror the range observed across mice, suggesting that the mechanisms involved in establishing VM-IAP methylation frequency distributions are conserved in ESCs. Viewed from the perspective of the LAB-KZFP model, ESCs are 'trapped' in the stochastic establishment phase. Indeed, many KZFPs are highly expressed in ESCs. The lack of methylation maintenance may reflect inhibition of the maintenance methyltransferase DNMT1 or active demethylation of VM-IAPs by TET methylcytosine dioxygenases. In support of the latter possibility, a study on TET enzyme activity in ESCs found that TET1 and TET2 bind a number of retrotransposon classes (de la Rica *et al.* 2016). Determining whether VM-IAP methylation levels in ESCs become mitotically heritable upon differentiation is a critical next step in understanding this process.

7.2.3 ZFPs as the link between metastable epialleles and transgenes

Previous work on the A^{vy} and $Axin^{Fu}$ alleles suggested that metastable epialleles and transgenes are regulated by similar mechanisms. They are both associated with variable DNA methylation and both exhibit parent-of-origin and genetic background effects (see section 1.4.4.2). These parallels are further highlighted in this thesis. The analysis described in Chapter 5 showed that VM-IAPs are also susceptible to maternal and zygotic genetic background effects, and demonstrated that a subset of VM-IAPs are targeted for methylation by strain-specific modifiers (SSMs) in a manner analogous to the methylation of the HRD transgene by SSM1B (or ZFP979) (Ratnam *et al.* 2014). Ratnam and colleagues alluded to preliminary ChIP-seq data that indicates that repetitive sequences, including SINEs, are endogenous targets of ZFP979. Unfortunately, this analysis has not been published yet. This finding would strongly suggest that KZFPs lie at the root of the epigenetic similarities between transgenes and VM-IAPs. It is possible that the 'foreignness' of transgenes and VM-IAPs triggers the same host defence mechanism. However, ZFP979-mediated repression of the HRD transgene strictly relies on binding at the *gpt* gene located within the HRD construct and derived from *E. coli*. A nearly identical transgene that lacks the *gpt* gene is not subject to ZFP979-dependent methylation (Engler *et al.* 1998). This indicates that ZFP979-induced methylation is highly sequence-specific and

argues against a general mechanism targeting foreign DNA indiscriminately. It follows that the *gpt* gene is similar in sequence to the TEs targeted endogenously by ZFP979.

Additional evidence for a KZFP-mediated mechanistic overlap between metastable epialleles and transgenes comes from the genome-wide mutagenesis screen for modifiers of transgene variegation conducted by Emma Whitelaw and colleagues (Blewitt *et al.* 2005; Daxinger *et al.* 2013). The screen identified more than 50 mutations that influence the maintenance of variegation of a GFP transgene, most of which occurred in genes with known roles in epigenetic regulation. A subset of these were tested for their effect on A^{vy} coat colour. Out of five yellow-shifting mutations, four map to genes involved in KZFP-mediated repression, including *Kap1*, *Setdb1*, *Hdac1*, and *Dnmt1* (Blewitt *et al.* 2005; Chong *et al.* 2007; Daxinger *et al.* 2016). The fifth yellow-shifting mutation was located in *Smchd1*. This gene was unknown prior to the screen and is now recognised as an important regulator of X-inactivation (Blewitt *et al.* 2005, 2008). While SMCHD1 has not been reported to interact with KZFPs, SMCHD1 occupancy overlaps with CTCF binding sites (Chen *et al.* 2015). Considering that VM-IAPs are enriched for CTCF (Kazachenka *et al.* 2018), it is possible that SMCHD1 contributes to VM-IAP methylation states by competing for binding with CTCF.

7.3 Mechanistic alternatives

The LAB-KZFP model should be considered alongside the following mechanistic nuances:

- The LAB-KZFP model suggests that the ancestral form of VM-IAP-modifying KZFPs bind their targets with high affinity and that low-affinity KZFPs are the product of mutations away from the ancestral form, resulting in strain-specific VM-IAP modification. However, it is also possible that high-affinity KZFPs evolved from low-affinity KZFPs. In the case of $Ssm_{Rab6b}{}^{C}$, this would have to be the product of convergent evolution because both the CAST and 129Sv mouse strains appear to have the high-affinity $Ssm_{Rab6b}{}^{C}$ allele, while the

B6 strain does not. Once *Ssm*$_{Rab6b}^{C}$ is identified, a cross-strain phylogenetic analysis on this locus will resolve this question.

- Low-affinity binding of KZFPs to VM-IAPs could be the product of recent spontaneous mutations but could also be symptomatic of strain-specific natural selection. If the latter were the case, low-affinity binding to VM-IAPs may be a transitory state reflecting the strain-specific evolution of KZFPs towards the effective targeting of a different class of repeat element or even of unique genomic regions.

- KZFP modifiers may bind to VM-IAPs with high affinity via their zinc finger array but recruit KAP1 with low affinity due to a mutation in their KRAB domain. This would result in similar outcomes but is less likely given that the evolutionary pressures exerted on KZFPs promote zinc finger binding diversification and KRAB domain conservation (Bellefroid *et al.* 1991).

- VM-IAP-modifying KZFPs may be expressed stochastically in early development. Under this premise, VM-IAP targets would be efficiently methylated in cells expressing their KZFP modifier(s) but unmethylated in non-expressing cells. Although stochastic gene expression has been reported in a number of experimental systems and model organisms (reviewed in Raj & van Oudenaarden 2008), the strain-specificity of VM-IAP metastability would likely require strain-specific regulators of stochastic gene expression.

- *Ssm*$_{Rab6b}^{C}$ may map to a different class of epigenetic regulator or to a previously uncharacterised gene. If the latter turns out to be the case, elucidating the function of the novel modifier will be an interesting avenue of future research.

- This discussion has largely focused on the potential role of KZFPs in the establishment of metastability at VM-IAPs, but it is conceivable that KZFPs also contribute to the maintenance of VM-IAP methylation states. This is illustrated by the murine ZFP57 and ZFP445 proteins, which act together to maintain germline-derived DNA methylation patterns at imprinted regions

during the post-fertilisation wave of global methylation erasure (Li *et al.* 2008; Strogantsev *et al.* 2015; Takahashi *et al.* 2019).

In addition to the mechanistic subtleties listed above, a fundamentally different framework can be envisioned that does not presume that VM-IAPs are actively targeted for methylation by a KZFP or another epigenetic modifier. In an alternative 'protection' model, VM-IAPs are variably protected from *de novo* methylation, much like the *Aprt* gene promoter is protected against methylation by the SP1 transcription factor (Brandeis *et al.* 1994; Macleod *et al.* 1994). According to this model, VM-IAPs are methylated by default and the modifiers that drive metastability either prevent methylation from occurring faithfully or actively contribute to VM-IAP demethylation. A study that compared methylated and unmethylated IAP LTRs found an enrichment in transcription factor binding sites in unmethylated LTRs, which led to the proposition that transcription factor occupancy interferes with methylation of these elements (Shimosuga *et al.* 2017). In the context of Ssm_{Rab6b}, a protection model would suggest that the Ssm_{Rab6b^B} allele in B6 mice codes for a protective modifier and that the Ssm_{Rab6b^C} allele in CAST mice is either non-functional or absent. However, the hypermethylation observed in F1 hybrids is strong evidence against this model because these individuals still carry a copy of the hypothetically protective Ssm_{Rab6b^B} allele. While it is possible that the mechanism is dosage-sensitive, requiring two copies of B6-derived Ssm_{Rab6b^B} to induce methylation variability, the simplest explanation is that the CAST-derived Ssm_{Rab6b^C} allele is a dominant repressor of VM-IAPs regardless of copy number. Thus, the targeted model is preferred over a protection model.

It is also possible that SSM-mediated hypermethylation of VM-IAPs is mechanistically distinct from the establishment of methylation stochasticity in the absence of the SSM. For example, if SSM_{Rab6b^C} were to map to a CAST-specific *Kzfp*, that would not necessarily implicate the corresponding B6 locus in the establishment of variable methylation at VM-IAP$_{Rab6b}$, VM-IAP$_{Pink1}$, and VM-IAP$_{Ect2l}$ in a B6 context. In other words, VM-IAPs may be targeted for complete repression by KZFPs in strains harbouring the relevant *Kzfp* gene(s), but exhibit metastability via KZFP-independent mechanisms in strains lacking the relevant KZFP modifier(s). RNA-mediated regulation is a candidate

alternative pathway given the important role of regulatory RNA in TE repression (section 1.3.3.1).

7.4 Future directions

In addition to mapping the *Ssm*$_{Rab6b}^{C}$ locus, a number of experiments can be conducted to test the LAB-KZFP model. KAP1 recruitment and deposition of H3K9me3 are hallmarks of KZFP-induced repression. If SSM$_{Rab6b}^{C}$ is a KZFP, then VM-IAP$_{Rab6b}$, VM-IAP$_{Pink1}$, and VM-IAP$_{Ect2l}$ should show greater KAP1 and H3K9me3 occupancy in BC and CB F1 hybrid ESCs compared to B6 ESCs. This can be tested via ChIP-qPCR using KAP1 and H3K9me3 antibodies (provided the VM-IAPs exhibit hypermethylation in BC and CB ESCs, which should be checked in the first instance). After the identification of *Ssm*$_{Rab6b}^{C}$, follow-up experiments should aim to understand the mechanism by which it induces VM-IAP methylation. Their design will depend on the nature of the identified gene and could include gene editing of important binding sites as well as *in vitro* binding assays and expression vector experiments. Computational simulations of theoretical binding interactions would also aid in developing a more detailed model linking binding kinetics to the methylation frequency distributions observed at specific VM-IAPs.

The experiments from Chapter 5 highlight the value of using genetically distinct inbred mouse strains to identify drivers of epigenetic metastability. This method could also be applied to uncover A^{vy} and $Axin^{Fu}$ SSMs. The experimental design would consist of breeding A^{vy} and $Axin^{Fu}$ mice with a range of inbred mouse strains and screening for breeding partners that produce litters that lack variable expressivity. Mouse strains that give rise to purely pseudoagouti pups in the case of A^{vy} or purely straight-tailed pups in the case of $Axin^{Fu}$ may carry one or more SSMs that target the A^{vy} and $Axin^{Fu}$ IAPs, respectively.

The power and generalisability of this study would be increased by expanding the analysis to incorporate other strains and additional repeat element classes. The first step would be to screen for variably methylated repeat elements in a panel of strains to generate a comprehensive list of metastable epialleles, classified based on their presence or absence in each strain. A side-by-

side analysis comparing KZFP content and diversity across strains would provide a useful dataset to cross-reference against the catalogue of metastable epialleles, particularly if *Ssm*$_{Rab6b}^{C}$ is identified as a KZFP gene. Modifier candidates could be identified based on whether a KZFP allele is shared among strains exhibiting hypermethylation at a given region and absent from strains showing metastability at the same locus. Unfortunately, this programme of work is currently unrealistic due to the lack of high-quality reference genomes. Although substantial progress was made with the recent publication of 16 inbred mouse strain genomes (Lilue *et al.* 2018), they still contain large gaps in repetitive regions containing TEs and KZFP clusters. Long-read sequencing technology is expected to vastly improve these genomes and their annotations in the coming years.

In the meantime, much remains to be explored with regards to VM-IAP modifiers in the B6 genome. The method currently used for measuring VM-IAP methylation levels (bisulphite pyrosequencing) requires that each VM-IAP be assayed separately. For this reason, only twelve VM-IAPs were focused on in this thesis rather than the entire set of ~50 validated VM-IAPs. The development of a multi-locus capture method that allows the quantification of methylation levels at all VM-IAPs at the same time would broaden the scope of questions that can be addressed experimentally. This type of technology would enable more sophisticated cross-loci analyses and could be used to assess some of the properties investigated in this dissertation (e.g. heritability and environmental susceptibility) in a more holistic fashion. It would also make possible the development of a screen for VM-IAP modifiers without having to select specific VM-IAPs *a priori*. The experiment could be conducted in ESCs using a CRISPR-Cas9 screening library with VM-IAP methylation as a readout.

7.5 Functional implications and evolutionary perspectives

The functional and evolutionary relevance of incomplete epigenetic silencing at metastable epialleles remains an important unanswered question. The relationship between IAP methylation and phenotypic outcome observed in A^{vy} and $Axin^{Fu}$ mice indicates that metastable epialleles can have a profound

influence on phenotype, potentially acted upon both positively and negatively by natural selection. However, a direct impact of VM-IAP methylation on LTR promoter activity that subsequently influences neighbouring gene expression is only rarely observed (Kazachenka *et al.* 2018). This shows that metastability *per se* is not maintained as a product of host genome hijacking of *cis* regulatory promoter sequences present in repeat elements. Retroelements have been reported to act as enhancers to regulate host gene expression in a longer range *cis* capacity in addition to their better-described promoter potential (Sundaram *et al.* 2014). This along with the recently reported enrichment of CTCF binding at VM-IAP flanking regions suggests that metastable epialleles may play a role in orchestrating long-range regulatory networks (Kazachenka *et al.* 2018). 4C-seq (Circularized Chromosome Conformation Capture-seq) interactomes for a subset of VM-IAPs have been generated by Amir Hay in our lab to address this question. This possibility has important phenotypic implications and could contribute to our understanding of mammalian gene regulation more generally, potentially providing mechanistic insight into variable penetrance and expressivity. The functional significance of VM-IAPs can also be interrogated via conservation-based studies examining whether strain-specific VM-IAPs contribute to phenotypic differences between inbred mouse strains.

The metastable epialleles A^{vy} and $Axin^{Fu}$ are associated with evolutionarily young IAPs carrying LTRs of the IAPLTR1_Mm subclass (Faulk *et al.* 2013a; Qin *et al.* 2010) and VM-IAPs are enriched for young IAP elements that show high levels of absence/presence polymorphism across mouse strains (Kazachenka *et al.* 2018). It is possible that inter-individual methylation variation reflects a transient epigenetic state associated with a recently inserted element that has not yet been fully repressed. This hypothesis suggests that metastability represents a snapshot in evolutionary time and a phenotypically inconsequential phenomenon. In the cases where the variably methylated TE has an effect on host genome function, selective pressures could stabilise the epigenetic state of the element to an optimal level, potentially in a tissue-specific manner (Reiss & Mager 2007). This possibility would allow for the stabilisation of the element to a hypomethylated state, which may represent the first step in the process of TE domestication.

Alternatively, locus-specific epigenetic variability itself may confer an evolutionary advantage. It has been proposed through mathematical modelling that stochastic methylation at repeat elements could allow for more rapid fixation of the element and its associated genes, as well as increase the probability of fixation in the first place (Branciamore *et al.* 2015). Other evolutionary models have also been proposed, whereby the phenotypic plasticity conferred by metastable epialleles enables rapid adaptation to sudden environmental changes (Sharif *et al.* 2013). Developmental epigenetic reprogramming of these loci is consistent with this theory, allowing the re-establishment of epigenetic marks according to new environmental cues and rendering the developing embryo responsive to the environment into which it will be born.

The question remains of whether metastability is symptomatic of incomplete silencing or whether selective pressures are maintaining epigenetic stochasticity at specific regions in the genome. Viewed from the perspective of the LAB-KZFP model, this question captures the difference between transient low-affinity KZFP modifiers whose VM-IAP binding residues will eventually evolve improved binding capabilities (or stop binding altogether in the case of eventual hypomethylation), versus positive selection of low-affinity modifiers to maintain inter-individual methylation variability. It is important to note that inbred laboratory mouse strains suffer from severe inbreeding depression, which may put unusual strain on host defence mechanisms against TE mobilisation and result in the appearance of VM-IAPs. Quantifying the presence and number of VM-IAPs in strains that are more closely related to their wild-derived counterparts would contribute to addressing this point.

New technologies and novel data analysis techniques have been and will continue to be essential for elucidating the prevalence, molecular drivers, and functional consequences of incomplete silencing of mammalian TEs. These include mathematical modelling and machine learning approaches, CRISPR-Cas genetic engineering, single-cell multi-omics, long-read sequencing, and optimisation of computational pipelines for the analysis of repetitive sequences. Combining these tools with comparative research across mouse strains as well as other model organisms and humans will enrich our understanding of the

evolutionary origins, conservation, and functional relevance of epigenetic metastability.

7.6 Concluding remarks

The multipronged characterisation of multiple VM-IAPs in this thesis re-contextualises metastable epialleles in the genomics era. This body of work challenges the general idea that epigenetic variability at metastable epialleles is accompanied by epigenetic inheritance and environmental sensitivity. It is now clear that these phenomena manifest in a locus- and context-specific manner. However, when they do arise, there is potential for metastable epialleles to regulate the genotype-to-phenotype process.

It is the predictable reconstruction of epigenetic variability across generations in the absence of genetic variation that truly sets metastable epialleles apart from other genomic regions. The value of studying these unique loci extends well beyond their possible functional importance. They can be used as models of biological stochasticity in early development, with the potential to provide mechanistic insight into other stochastic processes or serve as powerful tools for lineage tracing. In addition, the mapping of VM-IAP modifiers may uncover currently uncharacterised regulators of TE repression and epigenetic control more generally. The prospect of better understanding TE regulation has wide-ranging implications, from resolving the extent to which TEs have been co-opted by their eukaryotic hosts over evolution to treating human diseases caused by TE insertions.

Chapter 8

Methods

8.1 Mouse procedures

Mouse work was regulated under the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012 following ethical review by the University of Cambridge Animal Welfare and Ethical Review Body (Home Office project license numbers: PC9886123 and PC213320E). Mice were maintained and bred in temperature- and humidity-controlled conditions under a 12h light-dark cycle. Unless otherwise stated, mice were fed a standard chow diet *ad libitum.*

8.1.1 Breeding experiments

The inheritance study described in Chapter 2 was conducted using C57BL/6J mice. The reciprocal hybrid experiments described in Chapters 5 and 6 were carried out using C57BL/6J and CAST/EiJ mice. For both studies, mice were set up for breeding at 8 weeks of age and the F1 methylation level was quantified from ear notches sampled from 10-day old pups. Male and female offspring from multiple litters per breeding pair were analysed.

8.1.2 *A^{vy}* mice

B6.C3H-Avy/J mice were maintained at The Jackson Laboratory by David Schroeder and Jennifer Stauffer. Sperm and spleen samples were collected from nine adult males ranging in coat colour from yellow to pseudoagouti.

8.1.3 *Dnmt3C* knockout mice

The *Dnmt3C* knockout mouse line was generated using CRISPR-Cas9 gene editing technology as previously described (Barau *et al.* 2016). The mouse line was housed in the pathogen-free Animal Care Facility of the Institut Curie in Paris, France.

8.1.4 *$Mtrr^{gt}$* mice

The *$Mtrr^{gt}$* mouse line was generated by randomly inserting a gene-trap (gt) vector into the *Mtrr* gene locus as previously described (Padmanabhan *et al.* 2013). *$Mtrr^{+/+}$* mice exhibit phenotypic abnormalities so C57BL/6J mice were used as controls. *$Mtrr^{gt/gt}$* mice were produced via *$Mtrr^{gt/gt}$* intercrosses. *$Mtrr^{+/+}$* and *$Mtrr^{+/gt}$* mice were produced via *$Mtrr^{+/gt}$* intercrosses. Liver, whole brains, and sperm were collected from fertile adult males. Eight males (from 3-7 separate litters) were assessed per tissue type per genotype. Brain and liver tissues were collected from different mice. *$Mtrr^{gt}$* genotyping and phenotyping was performed as previously described (Padmanabhan *et al.* 2013, 2017). Phenotypically normal E10.5 embryos showed no congenital malformations. Severely affected E10.5 embryos displayed one or more congenital malformation (e.g., placental phenotypes, neural tube closure defects, pericardial edema, reversed heart looping, overall abnormal morphology, etc.). For each phenotype, 6-7 embryos from 3-4 litters were assessed.

8.1.5 BPA exposed mice

Experimental conditions for this mouse model have been described previously (Susiarjo *et al.* 2015). C57BL/6J females were assigned one of three modified AIN 93G diets: control (7% corn oil, Harlan Teklad), lower BPA dose (10 μg/kg/day, Harlan Teklad), or upper BPA dose (10 mg/kg/day, Harlan Teklad). Dietary intervention began two weeks prior to mating and continued throughout pregnancy and lactation. All F1 pups were weaned onto the control diet. F1 adult livers were dissected from one mouse per litter, with reported sample sizes representing the number of litters.

8.1.6 Diet-induced obese mice

Experimental conditions for this mouse model have been described previously (Samuelsson *et al.* 2008). C57BL/6J females were assigned one of two diets: standard RM1 chow (7% simple sugars, 3% fat, 50% polysaccharide, 15% protein [w/w], 10.74 kJ/g; Special Dietary Services, Witham, UK) or a semi-synthetic energy-rich highly palatable obesogenic diet (10% simple sugars, 20% animal lard, 28% polysaccharide, 23% protein [w/w], 28.43 kJ/g; Special Dietary Services, Witham, UK) supplemented with sweetened condensed milk (16% fat, 33% simple sugars, 15% protein, 13.7 kJ/g; Nestle,UK) and fortified with mineral and vitamin mix AIN93G. Dietary intervention began three weeks prior to a first pregnancy that was performed to prove fertility and establish good maternal care. One week after weaning of pups from the first pregnancy, dams were mated for the experimental pregnancy. Dams were maintained on the same experimental diet through both pregnancies and lactation. All F1 pups were weaned onto the standard RM1 chow diet. F1 adult livers were dissected from one mouse per litter, with reported sample sizes representing the number of litters.

8.1.7 Ageing mice

8.1.7.1 Longitudinal study

10 C57BL/6J males and 10 C57BL/6J females from 5 different litters were ear notched at 10 days, 48 weeks, 80 weeks, and 112 weeks. Mice died of natural causes or were humanely culled for health reasons.

8.1.7.2 Cohort-based study

C57BL/6J mice were housed in the Babraham Institute Biological Services Unit (BI BSU) under Opportunistic Pathogen Free conditions. Liver samples were collected from 25 6-month-old and 25 24-month-old males by members of the Corcoran group.

8.2 Isolation of biological samples

8.2.1 Somatic tissue dissection

All other somatic tissues were snap frozen in liquid nitrogen following dissection and stored at -80°C before use. Liver samples were pulverized in liquid nitrogen prior to DNA or RNA extraction. Ear notch samples were stored at -20°C.

8.2.2 Sperm isolation

Sperm samples used in Chapter 3 were isolated with Dr. Rahia Mashoodh using a swim-up technique. The cauda epididymis and vas deferens were dissected from adult fertile males and transferred to 2 ml of pre-heated modified HTF medium in a 50 mm plastic dish. After making incisions in the cauda epididymides, caudal fluid was released into the medium via gentle squeezing. The contents of the plastic dish were transferred to a clean 14 ml round bottom

centrifuge tube and 3 ml modified HTF medium was added prior to a 30 min incubation at 37°C. Sperm-containing medium was filtered using a 70 um mesh and transferred to a clean 14 ml round bottom centrifuge tube, followed by centrifugation at 300 x g for 10 min at room temperature. The supernatant was aspirated and 5 ml of modified HTF medium was added before incubating the tubes at a 45° angle for 30 min at 37°C to allow sperm to swim up. 4 ml of supernatant was collected and 5 ml modified HTF medium was added to the original tube for a second swim-up incubation. Four swim-up incubations were performed in total, pooling all the supernatant into a single 15 ml conical centrifuge tube that was centrifuged at 2,000 x g for 5 min before aspirating the supernatant. The sperm pellet was resuspended with 1 ml PBS + 0.1% PVA and transferred to a 1.5 ml tube before centrifuging again at 2,000 x g for 5 min. A second wash was performed with 1 ml lysis buffer (0.1% SDS, 0.5% Triton-X) for 10 min on ice before centrifuging at 2,000 x g for 5 min. A final wash and centrifuge were carried out with 1 ml PBS + 0.1% PVA, and the final pellet was resuspended in 1X PBS and stored at -80°C. Sperm samples used in Chapter 4 were isolated by Dr. Gina Blake as described in Blake 2019.

8.2.3 P10 male germ cells

P10 male germ cells from the *Dnmt3C* knockout mouse line were isolated and sorted by FACS by Dr. Joan Barau as previously described (Barau *et al.* 2016).

8.2.4 Blastocyst isolation

Adult C57BL/6J females were set up for breeding with C57BL/6J males and checked for vaginal plugs every morning thereafter. The uterus and oviducts of pregnant females were dissected three days after the identification of a vaginal plug and flushed with M2 medium (supplemented with 0.06 g/L potassium penicillin-G and 0.05 g/L streptomycin sulfate) using a 1 mL syringe and 27G. Blastocysts were located under the microscope and transferred to clean M2 medium using a mouth aspirator and a pulled capillary tube. Healthy embryos fully developed to the blastocyst stage were transferred once again to a petri dish

with clean M2 medium. Blastocysts from a single mother were pooled and stored in a 0.5 ml Eppendorf tube at -80°C.

8.2.5 GV oocyte isolation

12-week-old C57BL/6J females were injected interperitoneally with 5 IU of gonadotropin. 40-42 hours later, ovaries were dissected and cleaned from surrounding fat before placing in M2 medium (supplemented with 0.06 g/L potassium penicillin-G, 0.05 g/L streptomycin sulfate, and 240 uM dbcAMP). Ovary follicles were punctures and pulled outward to release the oocytes using two 27G needles attached to 1 ml syringes. GV oocytes were transferred to clean supplemented M2 medium using a mouth aspirator and a pulled capillary tube. Oocytes were cleaned and separated from their associated cumulus cells using a mouth aspirator and a pulled capillary tube. Healthy GC oocytes were pooled and stored in a 0.5 ml Eppendorf tube at -80°C. Pools of 100 GC oocytes were used for analysis.

8.2.6 E16.5 embryonic tail and placenta dissections

Adult C57BL/6J females were set up for breeding with C57BL/6J males and checked for vaginal plugs every morning thereafter. 16 days after the identification of a vaginal plug and following confirmation of pregnancy via visual assessment, females were culled via cervical dislocation. The uterus was excised from the abdominal cavity and individual embryos were removed and rinsed in cold 1x PBS. After removal of the yolk sac and a subsequent additional rinse in cold 1X PBS, embryonic tails and full placentas were flash frozen in liquid nitrogen and stored at -80°C.

8.2.7 Tumour samples

Ageing mice suspected of harbouring internal growths were culled via cervical dislocation and inspected for tumours. Samples of discovered tumours

were collected and flash frozen in liquid nitrogen. When possible, a healthy sample of the affected organ was also collected. Samples were stored at -80°C.

8.3 Molecular methods

8.3.1 DNA extraction

8.3.1.1 Genotyping

DNA was extracted from ear notches for genotyping purposes using the PCRBIO Rapid Extract lysis kit (PCR Biosystems). Briefly, ear notches were incubated in 4 µl of 5X Buffer A, 2 µl of 10X Buffer B, and 14 µl ddH_2O at 75°C for 5 min followed by a 10 min inactivation step at 95°C. The sample was diluted 1:10 with ddH_2O and 1 µl was used as template for PCR.

8.3.1.2 Somatic tissues and ES cells

Samples were RNase A-treated at 37°C for 60 min and digested overnight with Proteinase K in lysis buffer (10 mM EDTA, 150 mM NaCl, 10 mM Tris-HCl pH 8, 0.1% SDS) at 55°C. Genomic DNA (gDNA) was isolated using a standard phenol-chloroform extraction and ethanol precipitation. Briefly, 1x volume of phenol was added to Proteinase K-digested samples and rotated gently at room temperature for 20 min. Following centrifugation at 12,000 x g for 20 min, the upper aqueous phase was transferred to a clean tube and 1x volume of phenol:chloroform: isoamyl alcohol (25:24:1, v/v) was added. This mixture was rotated and centrifuged as above and the upper aqueous phase was transferred to a clean tube. 1x volume of chloroform was added and a final round of rotation and centrifugation was carried out. Glycogen (1 ul) and NaCl (final concentration 0.2 M) was added to the upper aqueous phase and DNA was precipitated with 2-2.5x volumes of ice cold 100% ethanol. After incubation at 4°C for 1-24 hours, DNA was pelleted by centrifugation at 13 rpm for 30 min at 4°C and the supernatant was removed. The DNA pellet was washed with 70%

ethanol, centrifuged at 13 rpm for 10 min at room temperature, and all ethanol was removed before resuspending in ddH_2O. Liver gDNA used in the BPA exposure study was extracted and purified using the QIAamp DNA Mini Kit (QIAGEN) following the manufacturer's protocol. DNA quality and concentration were assessed using the Biodrop μLITE spectrophotometer and by running 1 μl DNA on a 0.8% agarose gel. DNA was kept at 4°C for short-term storage and -20°C for long-term storage.

8.3.1.3 Sperm

Sperm gDNA was extracted using the standard phenol chloroform extraction described above, except that the Proteinase K digestion was carried out in equal volumes of Solution A (75 mM NaCl pH 8; 25 mM EDTA) and Solution B (10 mM Tris-HCl pH 8; 10 mM EDTA; 1% SDS; 80 mM DTT). Stored frozen sperm (suspended in 1X PBS) was thawed and centrifuged for 5 min at 4,000 x g. The PBS supernatant was removed before adding Solution A and B.

8.3.1.4 Oocytes and blastocysts

Oocytes or blastocysts were pooled as required in 0.5 ml tubes (100 oocytes per pool, 5-8 blastocysts per pool). 14 μl ddH_2O, 1 μl 2 mg/ml Carrier RNA (QIAGEN), 1 μl 10% SDS, and 1 μl 10 mg/ml Proteinase K were added to the samples. Samples were incubated for 1 hour at 37°C and for 15 min at 98°C. The resulting extracted DNA was immediately bisulphite-converted as described below.

8.3.2 Polymerase chain reaction (PCR) genotyping

PCR for genotyping purposes was carried out using the REDTaq ReadyMix PCR Reaction Mix (Sigma-Aldrich). 7.5 μl Readymix was combined with 1 μl 10 μM forward primer, 1 μl 10 μM reverse primer, 1 μl DNA template, and 4.5 μl ddH_2O. The PCR conditions were (1) 95°C for 4 min 30 s; (2) 94°C for 30 s, optimised T°C for 30 s, 72°C for 30 s, 40 cycles; (3) 72°C for 5 min. Amplified DNA was evaluated by agarose gel electrophoresis. Agarose was dissolved in

0.5X Tris-borate-EDTA (TBE) buffer and 5 µl/ 100 µl buffer Safeview Nucleic Acid stain (Company) was added before setting the gel. Fragments greater than 200 bp were run on a 1.5% agarose gel; fragments smaller than 200 bp were run on a 2% agarose gel. Gels were run in 0.5X TBE buffer at 120 volts and visualized/photographed using the U:Genius^3 system (Syngene). Genotyping primers for *C57iap1* allelic variants are listed in **Table 8.1**.

8.3.3 Bisulphite conversion

DNA was bisulphite-converted using the Imprint® DNA Modification Kit (Sigma-Aldrich) according to the manufacturer's instructions for the two-step modification procedure. Briefly, 1 µg genomic DNA was adjusted for volume to 24 µl with ddH_2O and combined with 1 µl Balance Solution. Following incubation at 37°C for 10 min, 125 µl of DNA modification solution containing Balance Solution was added to the sample before incubating at 65°C for 90 min. Bisulphite-converted DNA was purified by spin column. 300 µl of Capture Solution was added to the Spin Column before loading the sample and centrifuging at 12,000 x g for 20 sec. The column was washed with 200 µl Cleaning Solution, centrifuged at 12,000 x g for 20 sec, incubated with 50 µl Balance/Ethanol Wash Solution for 8 min, and centrifuged again at 12,000 x g for 20 sec. Two final washes and 12,000 x g centrifugations with 200 µl 90% Ethanol were carried out before eluting the DNA with 20 µl Elution Solution. The sample was diluted with 20 µl ddH_2O before storage at -20°C.

8.3.4 Pyrosequencing

CpG-site specific methylation was quantified via pyrosequencing. Pyrosequencing assays were designed using PyroMark Assay Design SW 2.0 software (QIAGEN). PCR and sequencing primers are listed in

Table 8.2. PCR was carried out on bisulphite converted DNA using biotinylated reverse primers. Each PCR reaction consisted of 15.4 µl ddH_2O, 2 µl 10X CoralLoad PCR Buffer (QIAGEN), 0.5 µl 10 mM dNTPs mix, 1 µl 10 µM forward and reverse primers, and 0.1 µl HotStarTaq DNA Polymerase

(QIAGEN). The PCR conditions were (1) 95°C for 3 min; (2) 94°C for 30 s, Optimised T°C for 30 s, 72°C for 55 s, 40 cycles; (3) 72°C for 5 min. Each reaction was run in triplicate unless stated otherwise. 10-20 µl of the PCR product was bound to Streptavidin Sepharose High Performance beads (GE healthcare) by shaking at room temperature in binding buffer (10 mM Tris-HCL pH7.6, 2M NaCl, 1 mM EDTA, 0.1% Tween-20) for at least 5 min at 1,400 rpm. The beads and attached biotinylated strands were purified by sequential washing in 70% Ethanol, denaturation solution (0.2M NaOH), and wash buffer (10 mM Tris-acetate, pH 7.6) using the PyroMark Q96 Vacuum Workstation (QIAGEN). The purified product was resuspended in annealing buffer (20 mM Tris-acetate pH7.6, 2 mM magnesium acetate) and incubated at 85°C for 4 min with the sequencing primer. Following a 5 min incubation at room temperature, pyrosequencing was carried out on the PyroMark Q96 MD pyrosequencer (QIAGEN) using PyroMark Gold Q96 Reagents (QIAGEN). Percent methylation at each CpG was quantified using Pyro Q-CpG 1.0.9 software (Biotage). Methylation values across triplicates were averaged and consistency was assessed by calculating standard deviations. When indicated, methylation levels were averaged across CpGs for each individual at each locus.

8.3.5 Low-input pyrosequencing

Bisulphite pyrosequencing of oocyte and blastocyst DNA was carried out as described above with the following modifications. Bisulphite converted DNA was eluted from the spin column in 16 µl ddH_2O and 5 µl was used as template for PCR. Two rounds of PCR were conducted. Non-biotinylated forward and reverse primers for VM-IAP$_{Fam78b}$, VM-IAP$_{Marveld2,}$ and IAP$_{Pgm1}$ were combined for the first PCR, using the same conditions as above except only 20 cycles. For the second PCR, biotinylated reverse primers were used and each region was amplified separately using the PCR conditions described above. 1 µl of the product from the first PCR was used as template for the second PCR.

8.3.6 RNA extraction and cDNA synthesis

Total RNA was extracted from 20-30 mg of tissue using the AllPrep DNA/RNA Mini Kit (QIAGEN) according to the manufacturer's instructions. Tissue was placed in 2 ml screw-cap tubes containing Lysing Matrix D Bulk (MP) and homogenized in Buffer RLT at 6,000 x g for 40 s using the MagNA Lyser (Roche). For liver samples, 50% ethanol was used in the precipitation step to maximize RNA yield, as per the manufacturer's recommendation. A DNase I digestion was performed on the RNeasy spin column membrane using the RNase-Free DNase Set (QIAGEN). RNA integrity was assessed by running 1 µl RNA on a 1% agarose gel and checking for the presence of 28S and 18S rRNA bands. cDNA synthesis was carried out on 2 µg RNA using the RevertAid H Minus First Strand cDNA Synthesis kit (Thermo Scientific) according to the manufacturer's instructions. Random hexamer primers were used for the reaction. Minus RT samples were generated for each RNA sample to control for gDNA contamination. Thermocycler conditions were (1) 25°C for 5 min; (2) 42°C for 60 min, 70°C for 5 min, 10°C forever. cDNA was diluted by a factor of 5 and stored at -20°C.

8.3.7 Reverse transcription PCR (RT-PCR)

RT-PCR primers were designed using Primer3 software and are listed in **Table 8.3**. Primers for housekeeping gene *Hprt1* were taken from Stephens *et al.* 2011. A DNA dilution series was performed for each primer pair to assess amplification efficiency. All RT-PCR reactions were carried out on the LightCycler 480 Instrument (Roche). Each reaction contained 5 µl Brilliant III Ultra-Fast SYBR® Green QPCR Master Mix (Agilent), 1 µl ddH_2O, 3 µl 1 µM forward and reverse primers, and 1 µl 1:5 cDNA. The PCR conditions consisted of pre-incubation at 95°C for 5 min followed by 45 cycles of 95°C for 10 s, 60°C for 10s, 72°C for 10s. A melting curve analysis of 65°C to 95°C was performed after amplification to assess product specificity. Minus RT controls were run for each sample and no-template controls were run for each primer pair. All reactions were performed in triplicate. Relative expression was normalised to *Hprt1* expression and calculated using the ΔCt method.

8.3.8 Mass spectrometry

C57BL/6J and *Mtrr*$^{gt/gt}$ liver DNA was degraded into individual nucleosides using DNA Degradase Plus (Zymo) according to the manufacturer's instructions. 100-200ng degraded DNA was sent to the Babraham Institute Mass Spectrometry Facility, where global cytosine (C) and 5-Methylcytosine (5mC) levels were determined by liquid chromatography-tandem mass spectrometry as previously described (Ficz *et al.* 2013). All samples and standards were spiked with stable-isotope-labelled internal standards: 2'-deoxycytidine-$^{13}C_1{}^{15}N_2$ (Santa Cruz), 5-(methyl-$^2H_3$)-2'-deoxycytidine (Santa Cruz), and 5-(hydroxymethyl-$^2H_3$)-2'-deoxycytidine (Toronto Research Chemicals). Global 5mC levels are reported as percentages relative to C.

8.3.9 Sanger sequencing

Three PCR primer pairs were designed to amplify IAP$_{Pgm1}$ prior to Sanger sequencing (**Table 8.1**). The primer pair named *IAP$_{Pgm1}$_5′ frag* was designed to amplify a 4,992 bp product with one primer inside the IAP and the other in the unique region of the 5′ border. The primer pair named *IAP$_{Pgm1}$_3′ frag* was designed to amplify a 4,343 bp product with one primer inside the IAP and the other in the unique region of the 3′ border. The primer pair named *IAP$_{Pgm1}$_full* was designed to amplify a 8,960 bp fragment (at least according to the GRC38/mm10 genome assembly) with both primers in the unique regions on either side of the IAP. PCR amplification of these fragments was carried out using the Expand Long Template PCR System (Roche) according to the manufacturer's instructions. Each PCR reaction contained 40 µl ddH_2O, 5 µl 10X PCR Buffer 2, 1.75 µl 10mM dNTPs, 1.5 µl 10 µM forward and reverse primers, and 1 µl enzyme mix. The PCR conditions were (1) 94°C for 2 min; (2) 94°C for 10 s, Optimised T°C for 30 s, 68°C for 6 min, 10 cycles; (3) 94°C for 15 s, Optimised T°C for 30 s, 68°C for 6 min + 20 s each successive cycle, 20 cycles; (4) 68°C for 7 min. A second nested PCR was performed using the product from the first PCR (diluted 1:20) as template. Each nested PCR reaction contained 15.2 µl ddH_2O, 2 µl CoralLoad PCR Buffer, 0.7 µl 10 mM dNTPs, 1 µl 10 µM forward and reverse nested primers, and 0.1 µl HotStarTaq DNA Polymerase (QIAGEN). The

primers used for the nested PCRs of the 5′ fragment, 3′ fragment, and entire IAP were designed to amplify 4,692 bp, 4,007 bp, and 8,516 bp fragments, respectively (**Table 8.1**). The conditions for the nested PCR were (1) 95°C for 3 min; (2) 94°C for 30 s, Optimised T°C for 30 s, 72°C for 55 s, 40 cycles; (3) 72°C for 5 min. PCR products were purified by gel extraction using the QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's instructions. DNA was eluted in 30 µl warm ddH_2O.

Sanger sequencing was carried out by Source BioScience. Samples were sent in a 5 µl volume and contained 10 ng/µl PCR product and 3.2pmol/µl sequencing primer. Sequencing primers were interspersed across the IAP element (**Table 8.4**). The quality of the sequence traces was visually examined and reliable sequences were merged using the EMBOSS *merger* tool.

8.3.10 Generation of embryonic stem cells (ESCs)

ESCs were generated with Dr. Nozomi Takahashi. Adult C57BL/6J females were set up for breeding with C57BL/6J males and checked for vaginal plugs in the morning. In the evening two days after the identification of a vaginal plug, embryos were flushed from the oviducts and placed in pre-equilibrated organ culture dishes containing KSOM+2i (KSOM medium + 0.2 µM MEK inhibitor PDO325901 + 3 µM GSEK3 inhibitor CHIR99021) drops covered with mineral oil. Embryos were incubated for one day at 37°C in a CO_2 incubator and transferred to a pre-equilibrated dish containing drops of N2B27+2i (N2B27 medium + 0.2 µM PDO325901 + 3 µM CHIR99021) + LIF. Embryos were incubated for two days at 37°C in a CO_2 incubator. Unhatched embryos were treated with pronase for 8-10 min at 37°C to remove the zona pellucida. Embryos were placed in a pre-equilibrated dish containing N2B27 +2i+LIF+ 20% anti-mouse serum and incubated for an hour at 37°C in a CO_2 incubator before rinsing 3 times in pre-equilibrated N2B27 +2i+LIF. Embryos were transferred to a pre-equilibrated dish of N2B27 + 20% freshly thawed rat serum (Sigma-Aldrich) and incubated at 37°C in a CO_2 incubator for 10 min. Each embryo was placed in a small drop of N2B27 +2i+LIF under oil and incubated for 30 min at 37°C in a CO_2 incubator. The trophectoderm lysates was removed using a pulled Pasteur pipette and placed in PCR tubes containing 10 µl PCR buffer with Proteinase K

(50 mM KCl, 10 mM TrisHCl pH 8.3, 2.5 mM MgCl2, 0.1 mg/mL gelatin, 0.45% NP40, 0.45% Tween 20, 200 μg /ml Pro K). Trophectoderm samples were incubated at 55°C for 1 hour and at 95°C for 10 min before storing at -20°C for later genotyping. Inner cell masses (ICMs) were isolated using pulled capillary tips and placed in a well of a laminin-coated 96-well plate pre-equilibrated with N2B27+2i+LIF and incubated at 37°C in a CO_2 incubator for 3 days, replacing half of the media on the second day. ICMs were disaggregated with a 10 min accutase treatment at 37°C in a CO_2 incubator followed by manual disaggregation using a capillary tip. All disaggregated cells were transferred to a new laminin-coated 96-well plate containing N2B27+2i+LIF (MEK inhibitor PDO325901 concentration increased to 1 μM) and incubated for 4 days at 37°C in a CO_2 incubator, replacing half of the media on the second day. To passage the cells, all media was removed and cells were rinsed with PBS before treating with accutase for 10 min at 37°C in a CO_2 incubator. Cells were manually dislodged and transferred to a 15 ml falcon tube before centrifuging for 3 min at 1, 000 rpm. The supernatant was removed and 500 μl N2B27+2i+LIF was added before placing the cells in a laminin-coated 24-well plate already containing 500 μl pre-equilibrated N2B27+2i+LIF. Cell were incubated at 37°C in a CO_2 incubator for 4 days, replacing half of the media on the second day. On the fourth day, cells were passaged again. Cells growing well after the second passage were frozen down in freezing medium (KO serum replacement + 2i + DMSO) and stored in a liquid nitrogen tank. The cell lines used for the results presented in **Figure 3.13** have the following identification numbers: line 1, ESC_6; line 2, ESC_2; line 3, ESC_3; line 4, ESC_9.

Trophectoderm samples were sex-genotyped via PCR using primers targeting the Y-linked *SRY* gene (**Table 8.1**). Each PCR reaction contained 6.7 μl ddH_2O, 1 μl CoralLoad PCR Buffer, 0.25 μl 10 mM dNTPs, 0.5 μl 10 μM forward primer, 0.5 10 μl 10 μM reverse primer, and 0.05 μl HotStarTaq DNA Polymerase (QIAGEN). PCR conditions were (1) 95°C for 3 min; (2) 94°C for 30 s, 56°C for 30 s, 72°C for 55 s, 40 cycles; (3) 72°C for 5 min. PCR products were run on a 1.5% agarose gel and male embryos were kept to generate ESC lines.

8.4 Computational and statistical analyses

8.4.1 Linear mixed-effects models

8.4.1.1 Inheritance study

The effect of parental methylation level on offspring methylation level was analysed using REML-fitted linear mixed-effects models (LMMs) in R via the lmer() function in the lme4 package (Bates et al., 2015). The bisulphite pyrosequencing methylation levels of all individuals used for this analysis were averaged across the first four CpGs of the IAP LTR and run through a logit transformation before feeding into the model. The LMMs for each VM-IAP included maternal methylation level, paternal methylation level, and sex as fixed effects. Breeding pair as well as litter nested within breeding pair were treated as random intercept effects, accounting for the nonindependence of siblings and littermates, respectively. Interaction between maternal and paternal methylation levels was originally assessed but no significant interaction was found. Parameter estimates, standard errors, and t values are reported in Figure S5B. By default, the reference intercept is selected alphabetically, in this case representing the estimate for female methylation level. To evaluate the significance of the fixed effects, p values were generated using the Satterthwaite approximation for degrees of freedom, applied by the lmerTest package in R (Kuznetsova et al., 2017). To account for multiple testing, q-values were generated using the Benjamini-Hochberg correction with a false discovery rate of 0.05 (Benjamini and Hochberg, 1995). To assess effect sizes, semi-partial R 2 values for each fixed effect were calculated using the r2beta() function from the r2glmm package in R (Jaeger et al., 2017). Parameter estimates, standard errors, t values, marginal R^2 values, and p-values are reported in **Table 2.2**.

8.4.1.2 Longitudinal ageing study

REML-fitted linear mixed-effects models (LMMs) were used to assess the effect of age on VM-IAP methylation levels. The analysis was carried in R using

the lmer() function in the lme4 package (Bates *et al.* 2015). The methylation levels of the first four CpGs at the VM-IAP LTRs were averaged for each individual at each time point. For VM-IAP$_{Rnf157}$ and VM-IAP$_{Rab6b}$, only CpGs 3 and 4 were averaged due to technical difficulties. All percentage methylation data were logit transformed prior to analysis. For each VM-IAP, a LLM was fitted with age as a fixed effect and litter and mouse ID as random intercept effects. Sex and mouse cage were explored as possible fixed and random effects, respectively, but both were excluded from the final models as doing so reduced the Akaike and Bayesian information criteria (AIC and BIC).

To evaluate effect sizes (i.e. the proportion of variance explained by age), marginal R^2 values were generated using the r2beta() function from the r2glmm package in R, applying the Nakagawa and Schielzeth method (Jaeger *et al.* 2017). P-values were calculated with the lmerTest package in R using the Satterthwaite approximation for degrees of freedom (Kuznetsova *et al.* 2017). To account for multiple testing, a Bonferroni-adjusted α value of 0.0056 was used to interpret p-values. Parameter estimates, standard errors, t values, marginal R^2 values, and p-values are reported in **Table 4.1**.

8.4.2 Co-variation analysis

Methylation levels at six VM-IAPs in 33 C57BL/6J mice were normalized to a given VM-IAP's inter-individual methylation range. A normalized correlation matrix showing Pearson correlation coefficients as well as p values for the correlation of each VM-IAP pair was generated using GraphPad Prism. A Bonferroni-adjusted a value of 0.008 was used.

8.4.3 Statistical tests

Unless otherwise stated, statistical analyses were carried out using GraphPad Prism 8 software. All statistical tests used are indicated in the appropriate figure legend. Welch's t-tests were applied when comparing methylation or expression levels between two groups. Paired data, such as matched embryonic tail and

placenta samples, were analysed using paired t-tests. Comparisons of three or more groups were analysed using Welch's ANOVA followed by Tamhane T2 or Dunnett's T3 post hoc tests, depending on whether all means were compared or each mean was compared to controls, respectively. Correlations between expression and methylation levels were tested by generating Pearson's correlation coefficients and two-tailed p-values. For the repeated inheritance experiment on VM-IAP$_{Gm13849}$, the mean methylation level of littermates was calculated for the first litter of five highly methylated and five lowly methylated mothers, and significance was assessed with a one-sided unpaired t test.

Histograms of B6, BC, and CB methylation levels were generated using 20 bins (width = 5%). The resulting distributions were compared using Kruskal-wallis tests. The analysis was repeated using only data collected from female BC and CB offspring to ensure genetic identity and the same effects were observed (data not shown). B6 frequency distributions were assessed for normality with Anderson-Darling, D'Agostino & Pearson, Shapiro-Wilk, and Kolmogorov-Smirnov tests (Appendix A, **Table A.3**). Data normality was also examined visually by generating QQ plots (Appendix A, **Figure A.1**). Chi-square analyses on backcrossed N2 methylation levels were conducted using the CHISQ.TEST function in Excel.

8.5 Primer sequences

Table 8.1: PCR primers

Locus	Forward	Reverse
VM-IAP$_{Rab6b}$_full	CTGTAGTTATGTTTTTATGTA	TACACCAAAGATCGCTGTCG
IAP$_{Pgm1}$_5′ frag[1]	ACTTCTCTAACAACGTTGAGCA	CAAAGGCCTTAACTCAGCGG
IAP$_{Pgm1}$_5′ frag_N[2]	GAGGGGCTAGGTGGGAATAC	CCTGTATCAGAAATTTGGAC**C**
IAP$_{Pgm1}$_3′ frag	AGTCCTACAATGTGCCAACTT	GCATTCAATGTTTACCTGCCAC
IAP$_{Pgm1}$_3′ frag_N	AGAAGCTCTTTTGCCAGTGAG	TTTCTCTTGGCGCACAGTTC
IAP$_{Pgm1}$_full	ACTTCTCTAACAACGTTGAGCA	GCATTCAATGTTTACCTGCCAC
IAP$_{Pgm1}$_full_N	GAGGGGCTAGGTGGGAATAC	TTTCTCTTGGCGCACAGTTC
C57iap1$^{full\ \&\ solo}$	CCAAGAGAAGAGAAATTACCTGGA	TCACGGGAAAGGTAGAGTAC
C57iap1^{full}	TTTCATGAAGGTTCAGTGTCCT	CAACAATCCATACTCAAGTGTCC
C57iap1^{solo}	CAATGGCATGTGCTTAGTGG	CAACAATCCATACTCAAGTGTCC
Sry	GCAGGCTGTAAAATGCCACT	TTCCAGGAGGCACAGAGATT

[1]frag, fragment
[2] N, nested

Table 8.2: Bisulphite pyrosequencing primers

Locus		Primer
VM-IAPs		
VM-IAP$_{Gm13710}$	PCR Fwd	GGAGGAAGTAATAGTGTTAATAGGT
	PCR Rev	[Btn]CTCCCTAATTAACTACAACCCATAAC
	Seq	AGTAATAGTGTTAATAGGTTAGA
VM-IAP$_{Mbnl1}$	PCR Fwd	AAGTTTGTAATGGTGGGAGATTAA
	PCR Rev	[Btn]CCCTCTTTAAAAAAAACAACCATTACCT
	Seq	AAAATAAATTGTGGGAAGT
VM-IAP$_{Bmf}$	PCR Fwd	TTTTGGTTGAAGGGATTTTTATAGT
	PCR Rev	[Btn]CTCCCTAATTAACTACAACCCATAAC
	Seq	TTTTATAGTTAATAATTGTTGGGA
VM-IAP$_{Slc15a2}$	PCR Fwd	GGAAGTATAGAGAGAGTTATGGGGTTTA
	PCR Rev	[Btn]CCTTCTATTTTAAAAAAAACAACCATTACC
	Seq	TTTTGGAGTGTGGGA
VM-IAP$_{Marveld2}$	PCR Fwd	AGGAAGGAAGGTGATATAGAGA
	PCR Rev	[Btn]CCCTAATTAACTACAACCCATAAC
	Seq	TTTAAGTTGGTAAATAAATAATTTG
VM-IAP$_{Diap3}$	PCR Fwd	AGATGTAAGATAGAAGGGGTTTT
	PCR Rev	[Btn]ACAAACTAAAAATTTCCCACTCTCC
	Seq	AGTTTTTGAAGATGTAAGTAATAAA
VM-IAP$_{Gm13849}$	PCR Fwd	AGGGAGTATTATTTTTTGATTGGTTGTAGT
	PCR Rev	[Btn]ACACAATTCTATTTCTAATCCATTATATCT
	Seq	TTTATTAATTTAGAATATAGGATGT
VM-IAP$_{Rab6b}$	PCR Fwd	AGTGGGTAGGGTAGGTTAGATAAA
	PCR Rev	[Btn]CCCTAATTAACTACAACCCATAACC
	Seq	ATGTAAATAGTTGTTGGGA
VM-IAP$_{Fam78b}$	PCR Fwd	TGTTTAAGTTGAGATGTGGATAGT
	PCR Rev	[Btn]ATCACTCCCTAATTAACTACAACC
	Seq	TTTATTTAGTATAATTTTGTTGGG
VM-IAP$_{Eps8l1}$	PCR Fwd	AAGGGGGATTGTGGTTGGTTATTAG
	PCR Rev	[Btn]TCACTCCCTAATTAACTACAACCCATAAC
	Seq	GGGTGGGTGTGGGAA

VM-IAP$_{Rnf157}$	PCR Fwd	TTTGGGATAGGATGAGAGTTAGTATTG
	PCR Rev	[Btn]ATCACTCCCTAATTAACTACAACC
	Seq	AGGAGGGTGTGGGAA
VM-IAP$_{Pink1}$	PCR Fwd	TTTGGGGTTGGGGGTTTG
	PCR Rev	[Btn]CACTTCCTTTCTTTCCCACTACCTTAT
	Seq	AGATTTTGGTTTGTTGTG
VM-IAP$_{Ect2l}$	PCR Fwd	GGGAAAGGTAGAGTATAAGTAGT
	PCR Rev	[Btn]CCACTACCCAACTAAATCATAACA
	Seq	TGTTTATTATTTAGAATATAGGATG
VM-IAP$_{Zfp619}$	PCR Fwd	AGTTTTATTAAAGGTGAAGAAGTGTAGA
	PCR Rev	[Btn]CTTCACTTAAAACCCATCACTCCCTAAT
	Seq	TGTAGATAGTTGTTGGGA
Avy_5′ LTR[1]	PCR Fwd	ATTTTTAGGAAAAGAGTAAGAAGTAAG
	PCR Rev	CTACAAAAACTCAAAAACTCA
	Seq	TAGAATATAGGATGTTAG

IAPs

IAPEz[2]	PCR Fwd	GAGGGTGGTTTTTTATTTTATGTGT
	PCR Rev	[Btn]ATCACTCCCTAATTAACTACAACC*
	Seq	TTTTTATTTTATGTGTTTTGTTTTT*
IAP$_{Cdc73}$	PCR Fwd	TGTGATAAAAGAAAATAGAGGGAATTAAG
	PCR Rev	[Btn]CCCTAATTAACTACAACCCATAACC
	Seq	TTTTATTTATAGTTTATATGTGGGA
IAP$_{H2T24}$	PCR Fwd	GGTAGAGTATATGGAGTGAAGAATTAT
	PCR Rev	[Btn]ATTCCCTAAAAAACAAACTCACTT
	Seq	TTGTTTATTATTTAGAATATAGTT
IAP$_{Diap3}$	PCR Fwd	GGGGTTTGATGTTTTATTTTATTTTGTTTA
	PCR Rev	[Btn]AACACATACCCAAATTATTTATTTACCACT
	Seq	GTGATGTTTTGTTGGGA
IAP$_{Dst}$	PCR Fwd	GATTAATGAAAGGATATAGGGAGAAAG
	PCR Rev	[Btn]AACACATACCCAAATTATTTATTTACCACT
	Seq	ATTTGTTGATTGTTGGGA
IAP$_{Cdk15}$	PCR Fwd	GAGTAAAGGTAGAGTATATGTAGTGG
	PCR Rev	[Btn]CACCAAAAAACAAAACCAAACTATACTT
	Seq	TGTTTATTATTTAGAATATAGGATG

IAP_{Ell2}	PCR Fwd	GGTTATGGAAGAAAATGTAAATGATAAAAG
	PCR Rev	[Btn]ACACAAATAATCATAAAATACCCTTAACA
	Seq	AAATTATTTTTATGTGTTGGGA
$IAP_{Slc24a2}$	PCR Fwd	GGGAAAAATAGAGTATAAGTGGT
	PCR Rev	[Btn]TAATCCTACTTCAACAACTAAACTATCTC
	Seq	GTTTATTAATTTAGAATATAGGATG
IAP_{Zak}	PCR Fwd	GTTGTTTTTTTGTTTTTTGGATGATTT
	PCR Rev	[Btn]TCCCTAATTAACTACAACCCATAACC
	Seq	TGTTTTTAGATATTGTGGGA

Imprinted regions

Peg3	PCR Fwd	TTGGATTGGTTAGAGAGGAAGT
	PCR Rev	[Btn]ACAATCTAATACACCCACACTAA
	Seq	GGAGAGATGTTTATTTTG
H19 ICR	PCR Fwd	GGGGGGTAGGATATATGTATTTTT
	PCR Rev	[Btn]ACCTCATAAAACCCATAACTATAAAATCAT
	Seq	GTGTGTAAAGATTAGGG
IG-DMR	PCR Fwd	GTGGTTTGTTATGGGTAAGTTT
	PCR Rev	[Btn]CCCTTCCCTCACTCCAAAAATTAA
	Seq	TGGTTTATTGTATATAATGT

IAP_{Pgm1}

IAP_{Pgm1}_5′ LTR	PCR Fwd	GTTTGGTTTTTATATAGAAGGAAAGAAGAA
	PCR Rev	[Btn]ACAAATAATCATAAAATACCCTTAACTCAT
	Seq	ATGTTTGTTTGTTGGG
IAP_{Pgm1}_5′ border_1	PCR Fwd	AGTAAGATAGTGGAATTATGAGAGAA
	PCR Rev	[Btn]ACATTTCTTTTCTTCTTTCCTTCTATAT
	Seq	ATAAAAATGAAGATGTATAATTAGA
IAP_{Pgm1}_5′ border_2	PCR Fwd	TTTTTATGGGTTTTGGGGATTAAATTGAAG
	PCR Rev	[Btn]CTCTAACCACACAAATACCCTAACATAAAC
	Seq	AGTAAGTGTTTTTAGGGT
IAP_{Pgm1}_5′ border_3,4	PCR Fwd	TTATAGATTTTTTATTGTAAAGGGGGTGGT
	PCR Rev	[Btn]CTCTAACCACACAAATACCCTAACATAAAC
	Seq	GGGGTGGTGTTAGTA

IAP$_{Pgm1}$_5′ border_5,6,7	PCR Fwd	AGAGTAAAGGAGGTTTGTAGAGTAT
	PCR Rev	[Btn]ACCCCTCAAATCTACATTTACATTCC
	Seq	GGTTTGTAGAGTATAGTTTAATG
IAP$_{Pgm1}$_5′ border_8,9	PCR Fwd	GGTGGAAAGTTTTTGTGTTAGGAAAATA
	PCR Rev	[Btn]CTATACTCTACAAACCTCCTTTACTC
	Seq	GGAAAATAGTGGAGGTTTTA
IAP$_{Pgm1}$_3′ LTR	PCR Fwd	TGGTTTTTAAAGATGTAAGTAATAAAGTT
	PCR Rev	[Btn]ACAACAATCCATACTCAAATATCC
	Seq	GTAGAAGATTTTGGTTTGT
IAP$_{Pgm1}$_3′ border_1,2,3	PCR Fwd	TGGGTATATTTTTAAATTGAGGTTGAGTT
	PCR Rev	[Btn]ATAACAAAAACCTAAAACAACTAATCC
	Seq	AGATGGTTATAGTTATTATGTG
IAP$_{Pgm1}$_3′ border_5,4	PCR Fwd	AGGGTAAAGTATTTGATAAGAGGAAGAG
	PCR Rev	[Btn]AAACCACAAAAAAACCAACTACTT
	Seq	ATTTGATAAGAGGAAGAGTT
IAP$_{Pgm1}$_3′ border_7,6	PCR Fwd	AGAAGTTGAAAAAGGTATTTAATGTTTATT
	PCR Rev	[Btn]AAAAACTCTTCCTCTTATCAAATACT
	Seq	TGTTATTTTTTTAATAAATTGATAG
IAP$_{Pgm1}$_3′ border_10,9,8	PCR Fwd	GTTATTGTTTTTATTGTTGGGATAAAGTAT
	PCR Rev	[Btn]CAATCTTCTATACCAAAAAATATCTCCT
	Seq	GTATAAGGTTAAGGTTGAGAA

[1]Faulk et al., 2013
[2]Barau et al., 2016

Table 8.3: qRT-PCR primers

Gene	Forward	Reverse
Hprt1[1]	GAGGAGTCCTGTTGATGTTGCCAG	GGCTGGCCTATAGGCTCATAGTGC
β-actin	TCTTTGCAGCTCCTTCGTTG	ATTCCCACCATCACACCCTG
Gm13710	TTTTCTTTGGACTGCGGCCT	TTCCCCAAACCACTCCCAAC
Slc15a2	TCCCCTGGATGTGAGGAATCT	GGTCTGCCTGATCTTCTCCG
Slc15a2_ex9-10	GCTTCAGGAACCGTTCTGAG	CGTCCATAATGAGGTGCTTTGG
Slc15a2_ex19-20	GCAGTGCATTGTGAAACGGA	CCCCGGTTGGTGATATTAGTG
Eps8l1	TGATCAACACCGTGGAACCA	AAGTCTGGGGAGGTAGGGTC
Eps8l1_exons1-2	GAGTCCTCAGGCACCTCC	ACTTTGCACTTGGTTTGGGT
Eps8l1_exons4-5	TGTCAATCACCTGGTCACCT	TCAGTAGCATCTCCTGTGCC
Bmf	GTCTCTTGACCGACAGGCAA	CAGGCAGAAGCTGGTCAGAA
Marveld2	ATGGGGTAGGTGTGGCAATG	ATAGATCCCCGTGTCCCACA
Rnf157	CAGGAAGTTCAGGTCGCTGT	CCAGCCACTTCATCATGGGA
Mbnl1	GCTTGCAAAAGTGGCACAGA	GCAGATTTGGCCCAATGGAG
Pgm1	GGGATGGTGGCTCTTCACTT	CCTTCCTTCAGGGCTATGGC
Rell1	ATAGCCCAGCCCTCTCTTCA	CCCTTGAGACTTCCGCTGTT
Tbc1d1	CCTTCTCACAGCCTGGACTG	TCCATGTCTGTGGGCTGAAC
5830416l119RIK	CTGGAGGCGGACATTTCACT	CACCTGCTTAGCTCTGCCTT

[1]Stephens et al., 2011

Table 8.4: Sanger sequencing primers

Primer ID	Sequence
IAP_{Pgm1}_S1	GAGCAAGAGCCAGCTGTTTT
IAP_{Pgm1}_S2	CAATGGCATGTGCTTAGTGG
IAP_{Pgm1}_S3	CCAAGAGAAGAGAAATTACCTGGA
IAP_{Pgm1}_S3.5	ACATTCGCCGTCACAAGAT
IAP_{Pgm1}_S4	CTCTCTCGCTGGCATCTCTC
IAP_{Pgm1}_S4.5	ATTGAAGGGGCTTTCGTTTT
IAP_{Pgm1}_S5	CCAGTGAGTACAGCTTTACGAGGT
IAP_{Pgm1}_S5.5	GGCCTTTCCAAGGGTCTT
IAP_{Pgm1}_S6	GAAGAAGCAGCCCATTACCA
IAP_{Pgm1}_S6.5	CTGCGTAGTGCGTCAGCAAT
IAP_{Pgm1}_S7	CGTAAATACGGAACCAATGC
IAP_{Pgm1}_S7.5	GCAGCTGCTTTGACTCCAG
IAP_{Pgm1}_S8	AGCTGCGCCTCTGATAGAAC
IAP_{Pgm1}_S8.5	GGAAAGCCTGGGCATTTTA
IAP_{Pgm1}_S9	ACCCAGGAAGCAGTCAGAGA
IAP_{Pgm1}_S9.5	GTCCTGTGCTCAAGCCCTAA
IAP_{Pgm1}_S10	TCCTTGATACCGGAGCAGAT
IAP_{Pgm1}_S10.5	AGGACAGCAAGGGAAATTCA
IAP_{Pgm1}_S11	TCCAGTGTGGGTTCCTCAAT
IAP_{Pgm1}_S11.5	GGGGTCTCCCTGTACTTTCC
IAP_{Pgm1}_S12	GAGGGAACAATTCCCCTCTT
IAP_{Pgm1}_S13	TGTGCCAACTTTATGTGCAAG
IAP_{Pgm1}_S14	TTTTCAAAAGCTGTTGGGAGAT
IAP_{Pgm1}_S15	TGGCAGCCACATCTAATGAT
IAP_{Pgm1}_S15.5	CCTCAAGTGGTAGAATGTTTAGTGG
IAP_{Pgm1}_S16	TAGAGCCCATTCAGGCCTAC
IAP_{Pgm1}_S16.5	AAAGCTGCTGTGAGTTCTTGC
IAP_{Pgm1}_S17	TCAAAAATTCCAACAGTTCTGC

IAP_{Pgm1}_S17.5	CCAGATAGGCCCAATGAGAT
IAP_{Pgm1}_S18	GTCCCTCGTCTTGGTGATGT
IAP_{Pgm1}_S18.5	GACAGCCTTGGCTCTGTCTG
IAP_{Pgm1}_S19	TGGATGGCTGAATTTGAACA
IAP_{Pgm1}_S19.5	AATAGGTCGCTGGCCACTC
IAP_{Pgm1}_S20	TTTCATGAAGGTTCAGTGTCCT
IAP_{Pgm1}_S20.5	TAATCTGCGCATGAGCCAAG
IAP_{Pgm1}_S21	TCTCTCGCTGGCATCTCTCT
IAP_{Pgm1}_S21.5	GGACACTTGAGTATGGATTGTTG
IAP_{Pgm1}_S22	CGGCAGCTATCAGAACACAA

Bibliography

Aekplakorn, W., Chailurkit, L., & Ongphiphadhanakul, B. (2015). Relationship of serum bisphenol A with diabetes in the Thai population, National Health Examination Survey IV, 2009. *Journal of Diabetes*, **7**(2), 240–249.

Ahmadkhaniha, R., Mansouri, M., Yunesian, M., Omidfar, K., Jeddi, M., Larijani, B., Mesdaghinia, A., & Rastkari, N. (2014). Association of urinary bisphenol a concentration with type-2 diabetes mellitus. *Journal of Environmental Health Science and Engineering*, **12**(1), 64.

Aktas, T. (2011). *MusD transposable elements and their impact on endogenous gene regulation* (Doctoral dissertation, Heidelberg University, Heidelberg, Germany). Retrieved from https://archiv.ub.uni-heidelberg.de/volltext server/12753/

Alfaradhi, M. Z., Kusinski, L. C., Fernandez-Twinn, D. S., Pantaleão, L. C., Carr, S. K., Ferland-McCollough, D., Yeo, G. S. H., Bushell, M., & Ozanne, S. E. (2016). Maternal Obesity in Pregnancy Developmentally Programs Adipose Tissue Inflammation in Young, Lean Male Mice Offspring. *Endocrinology*, **157**(11), 4246–4256.

Allen, N. D., Norris, M. L., & Surani, M. A. (1990). Epigenetic control of transgene expression and imprinting by genotype-specific modifiers. *Cell*, **61**(5), 853–861.

Andersen, E., Ingerslev, L. R., Fabre, O., Donkin, I., Altıntaş, A., Versteyhe, S., Bisgaard, T., Kristiansen, V. B., Simar, D., & Barrès, R. (2019). Preadipocytes from obese humans with type 2 diabetes are epigenetically reprogrammed at genes controlling adipose tissue function. *International Journal of Obesity*, **43**(2), 306–318.

Anderson, O. S., Nahar, M. S., Faulk, C., Jones, T. R., Liao, C., Kannan, K., Weinhouse, C., Rozek, L. S., & Dolinoy, D. C. (2012). Epigenetic responses following maternal dietary exposure to physiologically relevant levels of bisphenol A. *Environmental and Molecular Mutagenesis*, **53**(5), 334–342.

Aravin, A. A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K. F., Bestor, T., & Hannon, G. J. (2008). A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice. *Molecular Cell*, **31**(6), 785–799.

Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K., & Hannon, G. J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science*, **316**(5825), 744–747.

Argeson, A. C., Nelson, K. K., & Siracusa, L. D. (1996). Molecular basis of the pleiotropic phenotype of mice carrying the hypervariable yellow (Ahvy) mutation at the agouti locus. *Genetics*, **142**(2), 557–567.

Ashe, A., Morgan, D. K., Whitelaw, N. C., Bruxner, T. J., Vickaryous, N. K., Cox, L. L., Butterfield, N. C., Wicking, C., Blewitt, M. E., Wilkins, S. J., Anderson, G. J., Cox, T. C., & Whitelaw, E. (2008). A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development. *Genome Biol*, **9**(12), R182.

Badger, T. M., Ronis, M. J. J., Wolff, G., Stanley, S., Ferguson, M., Shankar, K., Simpson, P., & Jo, C.-H. (2008). Soy Protein Isolate Reduces Hepatosteatosis in Yellow A^{vy}/a Mice Without Altering Coat Color Phenotype. *Experimental Biology and Medicine*, **233**(10), 1242–1254.

Bansal, A., Li, C., Xin, F., Duemler, A., Li, W., Rashid, C., Bartolomei, M. S., & Simmons, R. A. (2019). Transgenerational effects of maternal bisphenol: a exposure on offspring metabolic health. *Journal of Developmental Origins of Health and Disease*, **10**(2), 164–175.

Bansal, A., Rashid, C., Xin, F., Li, C., Polyak, E., Duemler, A., van der Meer, T., Stefaniak, M., Wajid, S., Doliba, N., Bartolomei, M. S., & Simmons, R. A. (2017). Sex- and dose-specific effects of maternal bisphenol A exposure on pancreatic islets of first- and second-generation adult mice offspring. *Environmental Health Perspectives*, **125**(9), 097022.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., & Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science*, **354**(6314), 909–912.

Barde, I., Rauwel, B., Marin-Florez, R. M., Corsinotti, A., Laurenti, E., Verp, S., Offner, S., Marquis, J., Kapopoulou, A., Vanicek, J., & Trono, D. (2013). A KRAB/KAP1-miRNA cascade regulates erythropoiesis through stage-specific control of mitophagy. *Science*, **340**(6130), 350–353.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, **67**(1), 1–48.

Beamer, L. J. (2015). Mutations in hereditary phosphoglucomutase 1 deficiency map to key regions of enzyme structure and function. *Journal of Inherited Metabolic Disease*, **38**(2), 243–256.

Bell, A. C., & Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, **405**(6785), 482–485.

Bellefroid, E. J., Poncelet, D. A., Lecocq, P. J., Revelant, O., & Martial, J. A. (1991). The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **88**(9), 3608–3612.

Belshaw, R., Watson, J., Katzourakis, A., Howe, A., Woolven-Allen, J., Burt, A., & Tristem, M. (2007). Rate of recombinational deletion among human endogenous retroviruses. *Journal of Virology*, **81**(17), 9437–42.

Bernal, A. J., Dolinoy, D. C., Huang, D., Skaar, D. A., Weinhouse, C., & Jirtle, R. L. (2013). Adaptive radiation-induced epigenetic alterations mitigated by antioxidants. *The FASEB Journal*, **27**(2), 665–671.

Bertozzi, T. M., & Ferguson-Smith, A. C. (2019). Metastable epialleles and their contribution to epigenetic inheritance in mammals. *Seminars in Cell & Developmental Biology*. doi:10.1016/J.SEMCDB.2019.08.002

Bestor, T. H. (1990). DNA Methylation: Evolution of a Bacterial Immune Function into a Regulator of Gene Expression and Genome Structure in Higher Eukaryotes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **326**(1235), 179–187.

Bestor, T. H., Edwards, J. R., & Boulard, M. (2015). Notes on the role of dynamic DNA methylation in mammalian development. *Proceedings of the National Academy of Sciences*, **112**(22), 6796–6799.

Bestor, T., Laudano, A., Mattaliano, R., & Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. *Journal of Molecular Biology*, **203**(4), 971–983.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes & Development*, **16**(1), 6–21.

Bird, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Research*, **8**(7), 1499–1504.

Blake, G. E. T. (2019). *Investigating the Epigenetic Mechanism Behind Transgenerational Inheritance in Mice with Abnormal Folate Metabolism* (Doctoral dissertation, University of Cambridge, Cambridge, United Kingdom). Retrieved from https://www.repository.cam.ac.uk/handle/

1810/291811

Blewitt, M. E., Gendrel, A. V., Pang, Z., Sparrow, D. B., Whitelaw, N., Craig, J. M., Apedaile, A., Hilton, D. J., Dunwoodie, S. L., Brockdorff, N., Kay, G. F., & Whitelaw, E. (2008). SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nature Genetics*, **40**(5), 663–669.

Blewitt, M. E., Vickaryous, N. K., Hemley, S. J., Ashe, A., Bruxner, T. J., Preis, J. I., Arkell, R., & Whitelaw, E. (2005). An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc Natl Acad Sci U S A*, **102**(21), 7629–7634.

Blewitt, M. E., Vickaryous, N. K., Paldi, A., Koseki, H., & Whitelaw, E. (2006). Dynamic Reprogramming of DNA Methylation at an Epigenetically Sensitive Allele in Mice. *PLoS Genetics*, **2**(4), e49.

Blewitt, M., & Whitelaw, E. (2013). The use of mouse models to study epigenetics. *Cold Spring Harbor Perspectives in Biology*, **5**(11), a017939.

Bocklandt, S., Lin, W., Sehl, M. E., Sa, F. J., Sinsheimer, J. S., Horvath, S., & Vilain, E. (2011). Epigenetic Predictor of Age. *Plos ONE*, **6**(6), 1–6.

Boissinot, S., & Sookdeo, A. (2016). The Evolution of Line-1 in Vertebrates. *Genome Biology and Evolution*, **8**(12), 3485–3507.

Bostick, M., Kim, J. K., Esteve, P.-O., Clark, A., Pradhan, S., & Jacobsen, S. E. (2007). UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science*, **317**(5845), 1760–1764.

Bourc'his, D., & Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, **431**(7004), 96–99.

Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., & Bestor, T. H. (2001). Dnmt3L and the Establishment of Maternal Genomic Imprints. *Science*, **294**(5551), 2536–2539.

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten things you should know about transposable elements. *Genome Biology*, **19**(1), 199.

Branciamore, S., S. Rodin, A., Gogoshin, G., & D. Riggs, A. (2015). Epigenetics and Evolution: Transposons and the Stochastic Epigenetic Modification Model. *AIMS Genetics*, **2**(2), 148–162.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Names, A.,

Temper, V., Razin, A., & Cedar, H. (1994). Spl elements protect a CpG island from de novo methylation. *Nature*, **371**(6496), 435–438.

Brunmeir, R., Lagger, S., Simboeck, E., Sawicka, A., Egger, G., Hagelkruys, A., Zhang, Y., Matthias, P., Miller, W. J., & Seiser, C. (2010). Epigenetic Regulation of a Murine Retrotransposon by a Dual Histone Modification Mark. *PLoS Genetics*, **6**(4), e1000927.

Bultman, S. J., Klebig, M. L., Michaud, E. J., Sweet, H. O., Davisson, M. T., & Woychik, R. P. (1994). Molecular analysis of reverse mutations from nonagouti (a) to black-and-tan (at) and white-bellied agouti (AW) reveals alternative forms of agouti transcripts. *Genes and Development*, **8**(4), 481–490.

Carlson, L. L., Page, A. W., & Bestor, T. H. (1992). Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. *Genes & Development*, **6**(12B), 2536–41.

Carmell, M. A., Girard, A., van de Kant, H. J. G., Bourc'his, D., Bestor, T. H., de Rooij, D. G., & Hannon, G. J. (2007). MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline. *Developmental Cell*, **12**(4), 503–514.

Carone, B. R., Fauquier, L., Habib, N., Shea, J. M., Hart, C. E., Li, R., Bock, C., Li, C., Gu, H., Zamore, P. D., Meissner, A., Weng, Z., Hofmann, H. a., Friedman, N., & Rando, O. J. (2010). Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell*, **143**(7), 1084–1096.

Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., Duc, J., Jang, S. M., Turelli, P., & Trono, D. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes and Development*, **28**(13), 1397–1409.

Chaligné, R., & Heard, E. (2014). X-chromosome inactivation in development and cancer. *FEBS Letters*, **588**(15), 2514–2522.

Chang, B., Grau, T., Dangel, S., Hurd, R., Jurklies, B., Sener, E. C., Andreasson, S., Dollfus, H., Baumann, B., Bolz, S., Artemyev, N., Kohl, S., Heckenlively, J., & Wissinger, B. (2009). A homologous genetic basis of the murine cpfl1 mutant and human achromatopsia linked to mutations in the PDE6C gene. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(46), 19581–6.

Chen, K., Hu, J., Moore, D. L., … Blewitt, M. E. (2015). Genome-wide binding and mechanistic analyses of Smchd1-mediated epigenetic regulation. *Proceedings of the National Academy of Sciences of the United States of America*,

112(27), E3535-44.

Chong, S., Vickaryous, N., Ashe, A., Zamudio, N., Youngson, N., Hemley, S., Stopka, T., Skoultchi, A., Matthews, J., Scott, H. S., de Kretser, D., O'Bryan, M., Blewitt, M., & Whitelaw, E. (2007). Modifiers of epigenetic reprogramming show paternal effects in the mouse. *Nature Genetics*, **39**(5), 614–622.

Chung, B., Stadion, M., Schulz, N., Jain, D., Scherneck, S., Joost, H.-G., & Schürmann, A. (2015). The diabetes gene Zfp69 modulates hepatic insulin sensitivity in mice. *Diabetologia*, **58**(10), 2403–13.

Condic, M. L. (2015). The Role of Maternal-Effect Genes in Mammalian Development: Are Mammalian Embryos Really an Exception? *Stem Cell Reviews and Reports*, **12**(3), 276–284.

Cooney, C. A., Dave, A. A., & Wolff, G. L. (2002). Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring. *Journal of Nutrition*, **132**(8), 2393S–2400S.

Corsinotti, A., Kapopoulou, A., Gubelmann, C., Imbeault, M., Santoni de Sio, F. R., Rowe, H. M., Mouscaz, Y., Deplancke, B., & Trono, D. (2013). Global and Stage Specific Patterns of Krüppel-Associated-Box Zinc Finger Protein Gene Expression in Murine Early Embryonic Cells. *PLoS ONE*, **8**(2), e56721.

Cropley, J. E., Dang, T. H. Y., Martin, D. I. K., & Suter, C. M. (2012). The penetrance of an epigenetic trait in mice is progressively yet reversibly increased by selection and environment. *Proceedings of the Royal Society B: Biological Sciences*, **279**(1737), 2347–2353.

Cropley, J. E., Suter, C. M., Beckman, K. B., & Martin, D. I. K. (2006). Germ-line epigenetic modification of the murine A vy allele by nutritional supplementation. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(46), 17308–17312.

Cropley, J. E., Suter, C. M., Beckman, K. B., & Martin, D. I. K. (2010). CpG methylation of a silent controlling element in the murine A vy allele is incomplete and unresponsive to methyl donor supplementation. *PLoS ONE*, **5**(2), 1–7.

Cropley, J. E., Suter, C. M., & Martin, D. I. K. (2007). Methyl donors change the germline epigenetic state of the Avy allele. *The FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, **21**(12), 3021.

Czech, B., & Hannon, G. J. (2016). One Loop to Rule Them All: The Ping-Pong

Cycle and piRNA-Guided Silencing. *Trends in Biochemical Sciences*, **41**(4), 324–337.

Czeizel, A., Dudás, I., Vereczkey, A., Bánhidy, F., Czeizel, A. E., Dudás, I., Vereczkey, A., & Bánhidy, F. (2013). Folate Deficiency and Folic Acid Supplementation: The Prevention of Neural-Tube Defects and Congenital Heart Defects. *Nutrients*, **5**(11), 4760–4775.

Daujat, S., Weiss, T., Mohn, F., Lange, U. C., Ziegler-Birling, C., Zeissler, U., Lappe, M., Schübeler, D., Torres-Padilla, M.-E., & Schneider, R. (2009). H3K64 trimethylation marks heterochromatin and is dynamically remodeled during developmental reprogramming. *Nature Structural & Molecular Biology*, **16**(7), 777–781.

Daxinger, L., Harten, S. K., Oey, H., … Whitelaw, E. (2013). An ENU mutagenesis screen identifies novel and known genes involved in epigenetic processes in the mouse. *Genome Biology*, **14**(9), R96.

Daxinger, L., Oey, H., Apedaile, A., Sutton, J., Ashe, A., & Whitelaw, E. (2012). A forward genetic screen identifies eukaryotic translation initiation factor 3, subunit H (eIF3h), as an enhancer of variegation in the mouse. *G3 (Bethesda, Md.)*, **2**(11), 1393–6.

Daxinger, L., Oey, H., Isbel, L., Whitelaw, N. C., Youngson, N. A., Spurling, A., Vonk, K. K. D., & Whitelaw, E. (2016). Hypomethylation of ERVs in the sperm of mice haploinsufficient for the histone methyltransferase Setdb1 correlates with a paternal effect on phenotype. *Scientific Reports*, **6**, 25004.

De Bustros, A., Nelkin, B. D., Silverman, A., Ehrlich, G., Poiesz, B., & Baylin, S. B. (1988). The short arm of chromosome 11 is a "hot spot" for hypermethylation in human neoplasia. *Proceedings of the National Academy of Sciences of the United States of America*, **85**(15), 5693–5697.

de la Rica, L., Deniz, Ö., Cheng, K. C. L., Todd, C. D., Cruz, C., Houseley, J., & Branco, M. R. (2016). TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. *Genome Biology*, **17**(1), 234.

Deng, W., & Lin, H. (2002). miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Developmental Cell*, **2**(6), 819–830.

Dewannieux, M., Dupressoir, A., Harper, F., Pierron, G., & Heidmann, T. (2004). Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells. *Nature Genetics*, **36**(5), 534–539.

Dewannieux, M., & Heidmann, T. (2005). L1-mediated Retrotransposition of

Murine B1 and B2 SINEs Recapitulated in Cultured Cells. *Journal of Molecular Biology*, **349**(2), 241–247.

Dickies, M. M. (1962). A new viable yellow mutation in the house mouse. *The Journal of Heredity*, **53**, 84–6.

Dobie, K. W., Lee, M., Fantes, J. A., Graham, E., Clark, A. J., Springbett, A., Lathe, R., & McClenaghan, M. (1996). Variegated transgene expression in mouse mammary gland is determined by the transgene integration locus. *Proceedings of the National Academy of Sciences of the United States of America*, **93**(13), 6659–64.

Dokas, J., Chadt, A., Joost, H.-G., & Al-Hasani, H. (2016). Tbc1d1 deletion suppresses obesity in leptin-deficient mice. *International Journal of Obesity*, **40**(8), 1242–1249.

Dolinoy, D. C. (2008). The agouti mouse model: an epigenetic biosensor for nutritional and environmental alterations on the fetal epigenome. *Nutrition Reviews*, **66**, S7–S11.

Dolinoy, D. C., Huang, D., & Jirtle, R. L. (2007). Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proceedings of the National Academy of Sciences*, **104**(32), 13056–13061.

Dolinoy, D. C., Weidman, J. R., Waterland, R. a., & Jirtle, R. L. (2006). Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. *Environmental Health Perspectives*, **114**(4), 567–572.

Dolinoy, D. C., Weinhouse, C., Jones, T. R., Rozek, L. S., & Jirtle, R. L. (2010). Variable histone modifications at the A(vy) metastable epiallele. *Epigenetics : Official Journal of the DNA Methylation Society*, **5**(7), 637–44.

Doolittle, W. F., & Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature*, **284**(5757), 601–603.

Druker, R., Bruxner, T. J., Lehrbach, N. J., & Whitelaw, E. (2004). Complex patterns of transcription at the insertion site of a retrotransposon in the mouse. *Nucleic Acids Research*, **32**(19), 5800–8.

Ducker, G. S., & Rabinowitz, J. D. (2017). One-Carbon Metabolism in Health and Disease. *Cell Metabolism*, **25**(1), 27–42.

Duhl, D. M. J., Vrieling, H., Miller, K. A., Wolff, G. L., & Barsh, G. S. (1994). Neomorphic agouti mutations in obese yellow mice. *Nature Genetics*, **8**(1), 59.

Ecco, G., Cassano, M., Kauzlaric, A., Duc, J., Coluccio, A., Offner, S., Imbeault, M., Rowe, H. M., Turelli, P., & Trono, D. (2016). Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. *Developmental Cell*, **36**(6), 611–623.

Ecco, G., Imbeault, M., & Trono, D. (2017). KRAB zinc finger proteins. *Development*, **144**(15), 2719–2729.

Ehrlich, M., Gama-Sosa, M. A., Huang, L. H., Midgett, R. M., Kuo, K. C., Mccune, R. A., & Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Research*, **10**(8), 2709–2721.

Ekram, M. B., Kang, K., Kim, H., & Kim, J. (2012). Retrotransposons as a major source of epigenetic variations in the mammalian genome. *Epigenetics*, **7**(4), 370–382.

Elmore, C. L., Wu, X., Leclerc, D., Watson, E. D., Bottiglieri, T., Krupenko, N. I., Krupenko, S. A., Cross, J. C., Rozen, R., Gravel, R. A., & Matthews, R. G. (2007). Metabolic derangement of methionine and folate metabolism in mice deficient in methionine synthase reductase. *Molecular Genetics and Metabolism*, **91**(1), 85–97.

Elrod-Erickson, M., Benson, T. E., & Pabo, C. O. (1998). High-resolution structures of variant Zif268–DNA complexes: implications for understanding zinc finger–DNA recognition. *Structure*, **6**(4), 451–464.

Engler, P., Doglio, L. T., Bozek, G., & Storb, U. (1998). A cis-acting element that directs the activity of the murine methylation modifier locus Ssm1. *Proceedings of the National Academy of Sciences of the United States of America*, **95**(18), 10763–10768.

Engler, P., Haasch, D., Pinkert, C. A., Doglio, L., Glymour, M., Brinster, R., & Storb, U. (1991). A strain-specific modifier on mouse chromosome 4 controls the methylation of independent transgene loci. *Cell*, **65**(6), 939–947.

Engler, P., & Storb, U. (1987). High-frequency deletional rearrangement of immunoglobulin kappa gene segments introduced into a pre-B-cell line. *Proceedings of the National Academy of Sciences of the United States of America*, **84**(14), 4949–4953.

Faulk, C., Barks, A., & Dolinoy, D. C. (2013a). Phylogenetic and DNA methylation analysis reveal novel regions of variable methylation in the mouse IAP class of transposons. *BMC Genomics*, **14**(1), 48.

Faulk, C., Barks, A., Liu, K., Goodrich, J. M. J., & Dolinoy, D. C. D. (2013b). Early-

life lead exposure results in dose-and sex-specific effects on weight and epigenetic gene regulation in weanling mice. *Epigenomics*, **5**(5), 487–500.

Faulk, C., Liu, K., Barks, A., Goodrich, J. M., & Dolinoy, D. C. (2014). Longitudinal epigenetic drift in mice perinatally exposed to lead. *Epigenetics*, **9**(7), 934–941.

Feinberg, A. P., & Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*, **301**(5895), 89–92.

Feng, S., Cokus, S. J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M. G., Hetzel, J., Jain, J., Strauss, S. H., Halpern, M. E., Ukomadu, C., Sadler, K. C., Pradhan, S., Pellegrini, M., & Jacobsen, S. E. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(19), 8689–94.

Ferguson-Smith, A. C. (2011). Genomic imprinting: The emergence of an epigenetic paradigm. *Nature Reviews Genetics*, **12**(8), 565–575.

Ferguson-Smith, A. C., & Patti, M. E. (2011). You are what your dad ate. *Cell Metabolism*, **13**(2), 115–117.

Fernandez-Gonzalez, R., Ramirez, M. A., Pericuesta, E., Calle, A., & Gutierrez-Adan, A. (2010). Histone Modifications at the Blastocyst Axin1Fu Locus Mark the Heritability of In Vitro Culture-Induced Epigenetic Alterations in Mice. *Biology of Reproduction*, **83**(5), 720–727.

Festenstein, R., Tolaini, M., Corbella, P., Mamalaki, C., Parrington, J., Fox, M., Miliou, A., Jones, M., & Kioussis, D. (1996). Locus control region function and heterochromatin-induced position effect variegation. *Science*, **271**(5252), 1123–5.

Ficz, G., Hore, T. A., Santos, F., Lee, H. J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.-L., Walter, J., Cook, S. J., Andrews, S., Branco, M. R., & Reik, W. (2013). FGF Signaling Inhibition in ESCs Drives Rapid Genome-wide Demethylation to the Epigenetic Ground State of Pluripotency. *Cell Stem Cell*, **13**(3), 351–359.

Fort, A., Hashimoto, K., Yamada, D., … Carninci, P. (2014). Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nature Genetics*, **46**(6), 558–566.

Friedli, M., Nikolaev, S., Lyle, R., Arcangeli, M., Duboule, D., Spitz, F., & Antonarakis, S. E. (2008). Characterization of mouse Dactylaplasia

mutations: a model for human ectrodactyly SHFM3. *Mammalian Genome : Official Journal of the International Mammalian Genome Society*, **19**(4), 272–8.

Friedli, M., & Trono, D. (2015). The Developmental Control of Transposable Elements and the Evolution of Higher Species. *Annual Review of Cell and Developmental Biology*, **31**(1), 429–51.

Friso, S., Udali, S., De Santis, D., & Choi, S.-W. (2017). One-carbon metabolism and epigenetics. *Molecular Aspects of Medicine*, **54**, 28–36.

Gagnier, L., Belancio, V. P., & Mager, D. L. (2019). Mouse germ line mutations due to retrotransposon insertions. *Mobile DNA*, **10**(1), 15.

Garg, P., Joshi, R. S., Watson, C., & Sharp, A. J. (2018). A survey of inter-individual variation in DNA methylation identifies environmentally responsive co-regulated networks of epigenetic variation in the human genome. *PLOS Genetics*, **14**(10), e1007707.

Garrick, D., Fiering, S., Martin, D. I. K., & Whitelaw, E. (1998). Repeat-induced gene silencing in mammals. *Nature Genetics*, **18**(1), 56–59.

Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., Liu, C., Shi, W., & Bryant, S. H. (2009). The NCBI BioSystems database. *Nucleic Acids Research*, **38**(SUPPL.1), D492–D496.

Gelfman, S., Cohen, N., Yearim, A., & Ast, G. (2013). DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. *Genome Research*, **23**(5), 789–99.

Greenberg, M. V. C., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology*, 1–18.

Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F., Liu, W., Xie, Z. G., Shi, L., He, X., Jin, S. G., Iqbal, K., Shi, Y. G., Deng, Z., Szabó, P. E., Pfeifer, G. P., Li, J., & Xu, G. L. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, **477**(7366), 606–612.

Guilmatre, A., & Sharp, A. (2012). Parent of origin effects. *Clinical Genetics*, **81**(3), 201–209.

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.-P., Hu, B., Walsh, C. P., Li, J., Tang, F., & Xu, G.-L. (2014). Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote. *Cell Stem Cell*, **15**(4), 447–459.

Hackett, J. A., Sengupta, R., Zylicz, J. J., Murakami, K., Lee, C., Down, T. A., &

Surani, M. A. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science*, **339**(6118), 448–52.

Hackett, J. A., & Surani, M. A. (2013). DNA methylation dynamics during the mammalian life cycle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **368**(1609), 20110328.

Hadchouel, M., Farza, H., Simon, D., Tiollais, P., & Pourcel, C. (1987). Maternal inhibition of hepatitis B surface antigen gene expression in transgenic mice correlates with de novo methylation. *Nature*, **329**(6138), 454–456.

Haeussler, M., Zweig, A. S., Tyner, C., Speir, M. L., Rosenbloom, K. R., Raney, B. J., Lee, C. M., Lee, B. T., Hinrichs, A. S., Gonzalez, J. N., Gibson, D., Diekhans, M., Clawson, H., Casper, J., Barber, G. P., Haussler, D., Kuhn, R. M., & Kent, W. J. (2019). The UCSC Genome Browser database: 2019 update. *Nucleic Acids Research*, **47**(D1), D853–D858.

Hancks, D. C., & Kazazian, H. H. (2016). Roles for retrotransposon insertions in human disease. *Mobile DNA*, **7**(1), 9.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J. B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T., & Zhang, K. (2013). Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Molecular Cell*, **49**(2), 359–367.

Harris, R. A., Nagy-Szakal, D., & Kellermayer, R. (2013). Human metastable epiallele candidates link to common disorders. *Epigenetics : Official Journal of the DNA Methylation Society*, **8**(2), 157–63.

Harten, S. K., Bruxner, T. J., Bharti, V., Blewitt, M., Nguyen, T. M. T., Whitelaw, E., & Epp, T. (2014). The first mouse mutants of D14Abb1e (Fam208a) show that it is critical for early development. *Mammalian Genome*, **25**(7–8), 293–303.

Harten, S. K., Oey, H., Bourke, L. M., Bharti, V., Isbel, L., Daxinger, L., Faou, P., Robertson, N., Matthews, J. M., & Whitelaw, E. (2015). The recently identified modifier of murine metastable epialleles, Rearranged L-Myc Fusion, is involved in maintaining epigenetic marks at CpG island shores and enhancers. *BMC Biology*, **13**(1).

Heard, E., & Martienssen, R. A. (2014). Transgenerational Epigenetic Inheritance: Myths and Mechanisms. *Cell*, **157**(1), 95–109.

Hollick, J. B., Patterson, G. I., Coe, E. H., Cone, K. C., & Chandler, V. L. (1995). Allelic interactions heritably alter the activity of a metastable maize pl allele. *Genetics*, **141**(2), 709–719.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biol*, **14**(10), R115.

Hutnick, L. K., Huang, X., Loo, T.-C., Ma, Z., & Fan, G. (2010). Repression of Retrotransposal Elements in Mouse Embryonic Stem Cells Is Primarily Mediated by a DNA Methylation-independent Mechanism. *Journal of Biological Chemistry*, **285**(27), 21082–21091.

Imbeault, M., Helleboid, P. Y., & Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature*, **543**(7646), 550–554.

Isbel, L., Prokopuk, L., Wu, H., Daxinger, L., Oey, H., Spurling, A., Lawther, A. J., Hale, M. W., & Whitelaw, E. (2016). Wiz binds active promoters and CTCF-binding sites and is required for normal behaviour in the mouse. *eLife*, **5**, e15082.

Ishihara, H., Tanaka, I., Furuse, M., & Tsuneoka, K. (2000). Increased Expression of Intracisternal A-Particle RNA in Regenerated Myeloid Cells after X Irradiation in C3H/He Inbred Mice. *Radiation Research*, **153**(4), 392–397.

Jacobs, F. M. J., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A. D., Katzman, S., Paten, B., Salama, S. R., & Haussler, D. (2014). An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature*, **516**(7530), 242–245.

Jaeger, B. C., Edwards, L. J., Das, K., & Sen, P. K. (2017). An R2 statistic for fixed effects in the generalized linear mixed model. *Journal of Applied Statistics*, **44**(6), 1086–1105.

Jaenisch, R., Schnieke, A., & Harbers, K. (1985). Treatment of mice with 5-azacytidine efficiently activates silent retroviral genomes in different tissues. *Proceedings of the National Academy of Sciences of the United States of America*, **82**(5), 1451–1455.

Jain, D., Meydan, C., Lange, J., Claeys Bouuaert, C., Lailler, N., Mason, C. E., Anderson, K. V., & Keeney, S. (2017). rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. *PLOS Genetics*, **13**(8), e1006964.

Jansz, N., Keniry, A., Trussart, M., Bildsoe, H., Beck, T., Tonks, I. D., Mould, A. W., Hickey, P., Breslin, K., Iminitoff, M., Ritchie, M. E., McGlinn, E., Kay, G. F., Murphy, J. M., & Blewitt, M. E. (2018). Smchd1 regulates long-range chromatin interactions on the inactive X chromosome and at Hox clusters. *Nature Structural and Molecular Biology*, **25**(9), 766–777.

Jedrusik, A. (2015). Making the first decision: lessons from the mouse. *Reproductive Medicine and Biology*, **14**(4), 135–150.

Jern, P., & Coffin, J. M. (2008). Effects of Retroviruses on Host Genome Function. *Annual Review of Genetics*, **42**(1), 709–732.

Jirtle, R. L. (2014). The Agouti mouse: A biosensor for environmental epigenomics studies investigating the developmental origins of health and disease. *Epigenomics*, **6**(5), 447–450.

Johnson, K. R., Lane, P. W., Ward-Bailey, P., & Davisson, M. T. (1995). Mapping the mouse dactylaplasia mutation, Dac, and a gene that controls its expression, mdac. *Genomics*, **29**(2), 457–464.

Kaminen-Ahola, N., Ahola, A., Maga, M., Mallitt, K., Fahey, P., Cox, T. C., Whitelaw, E., & Chong, S. (2010). Maternal Ethanol Consumption Alters the Epigenotype and the Phenotype of Offspring in a Mouse Model. *PLoS Genetics*, **6**(1), e1000811.

Kanatsu-Shinohara, M., Takashima, S., Ishii, K., & Shinohara, T. (2011). Dynamic changes in EPCAM expression during spermatogonial stem cell differentiation in the mouse testis. *PloS One*, **6**(8), e23663.

Kano, H., Kurahashi, H., & Toda, T. (2007). Genetically regulated epigenetic transcriptional activation of retrotransposon insertion confers mouse dactylaplasia phenotype. *Proceedings of the National Academy of Sciences*, **104**(48), 19034–19039.

Karimi, M. M., Goyal, P., Maksakova, I. A., Bilenky, M., Leung, D., Tang, J. X., Shinkai, Y., Mager, D. L., Jones, S., Hirst, M., & Lorincz, M. C. (2011). Cell Stem Cell Article DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. *Stem Cell*, **8**(6), 676–687.

Kauzlaric, A., Ecco, G., Cassano, M., Duc, J., Imbeault, M., & Trono, D. (2017). The mouse genome displays highly dynamic populations of KRAB-zinc finger protein genes and related genetic units. *PloS One*, **12**(3), e0173746.

Kazachenka, A. (2018). *Identification and characterisation of murine metastable epialleles conferred by endogenous retroviruses* (Doctoral dissertation, University of Cambridge, Cambridge, United Kingdom). Retrieved from https://www.repository.cam.ac.uk/bitstream/handle/1810/276420/Kazachenka-2018-PhD.pdf?sequence=5&isAllowed=y

Kazachenka, A., Bertozzi, T. M., Sjoberg-Herrera, M. K., Walker, N., Gardner, J., Gunning, R., Pahita, E., Adams, S., Adams, D., & Ferguson-Smith, A. C.

(2018). Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. *Cell*, **175**(5), 1259–1271.

Kearns, M., Preis, J., Mcdonald, M., Morris, C., & Whitelaw, E. (2000). Complex patterns of inheritance of an imprinted murine transgene suggest incomplete germline erasure. *Nucleic Acids Research*, **28**(17), 3301–3309.

Kent, W. J. (2002). BLAT--the BLAST-like alignment tool. *Genome Research*, **12**(4), 656–64.

Kessler, N. J., Waterland, R. A., Prentice, A. M., & Silver, M. J. (2018). Establishment of environmentally sensitive DNA methylation states in the very early human embryo. *Science Advances*, **4**(7), eaat2624.

Kigami, D., Minami, N., Takayama, H., & Imai, H. (2003). MuERV-L Is One of the Earliest Transcribed Genes in Mouse One-Cell Embryos1. *Biology of Reproduction*, **68**(2), 651–654.

Kjer-Nielsen, L., Holmber, K., Perera, J. D., & McCluskey, J. (1992). Impaired expression of chimaeric major histocompatibility complex transgenes associated with plasmid sequences. *Transgenic Research*, **1**(4), 182–187.

Klutstein, M., Nejman, D., Greenfield, R., & Cedar, H. (2016). DNA Methylation in Cancer and Aging. *Cancer Research*, **76**(12), 3446–3450.

Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y., Suzuki, Y., & Kono, T. (2012). Contribution of Intragenic DNA Methylation in Mouse Gametic DNA Methylomes to Establish Oocyte-Specific Heritable Marks. *PLoS Genetics*, **8**(1), e1002440.

Kuff, E. L., & Lueders, K. K. (1988). The Intracisternal A-Particle Gene Family: Structure and Functional Aspects. In *Advances in Cancer Research*, Vol. 51, Academic Press, pp. 183–276.

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T. W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., Lin, H., Matsuda, Y., & Nakano, T. (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development (Cambridge, England)*, **131**(4), 839–49.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T. W., Hata, K., Li, E., Matsuda, Y., Kimura, T., Okabe, M., Sakaki, Y., Sasaki, H., & Nakano, T. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes and*

Development, **22**(7), 908–917.

Kuroiwa, Y., Kaneko-Ishino, T., Kagitani, F., Kohda, T., Li, L.-L., Tada, M., Suzuki, R., Yokoyama, M., Shiroishi, T., Wakana, S., Barton, S. C., Ishino, F., & Surani, M. A. (1996). Peg3 imprinted gene on proximal chromosome 7 encodes for a zinc finger protein. *Nature Genetics*, **12**(2), 186–190.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, **82**(13), 1–26.

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., & Reik, W. (2003). Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis*, **35**(2), 88–93.

Latham, K. E., & Sapienza, C. (1998). Localization of genes encoding egg modifiers of paternal genome function to mouse chromosomes one and two. *Development*, **125**(5), 929–935.

Lavebratt, C., Almgren, M., & Ekström, T. J. (2012). Epigenetic regulation in obesity. *International Journal of Obesity*, **36**(6), 757–765.

Lee, J., Matsuzawa, A., Shiura, H., Sutani, A., & Ishino, F. (2018). Preferable in vitro condition for maintaining faithful DNA methylation imprinting in mouse embryonic stem cells. *Genes to Cells*, **23**(3), 146–160.

Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K., & Wutz, A. (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes & Development*, **24**(3), 265–76.

Li, E., Bestor, T. H., & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**(6), 915–26.

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., & Ferguson-Smith, A. C. (2008). A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints. *Developmental Cell*, **15**(4), 547–557.

Lilue, J., Doran, A. G., Fiddes, I. T., … Keane, T. M. (2018). Sixteen diverse laboratory mouse reference genomes define strain-specific haplotypes and novel functional loci. *Nature Genetics*, **50**(11), 1574–1583.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**(7271), 315–322.

Loche, E., Blackmore, H. L., Carpenter, A. A., Beeson, J. H., Pinnock, A., Ashmore, T. J., Aiken, C. E., de Almeida-Faria, J., Schoonejans, J. M., Giussani, D. A., Fernandez-Twinn, D. S., & Ozanne, S. E. (2018). Maternal diet-induced obesity programmes cardiac dysfunction in male mice independently of post-weaning diet. *Cardiovascular Research*, **114**(10), 1372–1384.

Lueders, K. K., Frankel, W. N., Mietz, J. A., & Kuff, E. L. (1993). Genomic mapping of intracisternal A-particle proviral elements. *Mammalian Genome*, **4**(2), 69–77.

Lukic, S., Nicolas, J.-C., & Levine, A. J. (2014). The diversity of zinc-finger genes on human chromosome 19 provides an evolutionary mechanism for defense against inherited endogenous retroviruses. *Cell Death & Differentiation*, **21**(3), 381–387.

Lupo, A., Cesaro, E., Montano, G., Izzo, P., & Costanzo, P. (2011). ZNF224: Structure and role of a multifunctional KRAB-ZFP protein. *The International Journal of Biochemistry & Cell Biology*, **43**(4), 470–473.

Lyon, M. F. (1961). Gene action in the X-chromosome of the mouse (mus musculus L.). *Nature*, **190**(4773), 372–373.

Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., Firth, A., Singer, O., Trono, D., & Pfaff, S. L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*, **487**(7405), 57–63.

Macleod, D., Charlton, J., Mullins, J., & Bird, A. P. (1994). Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes & Development*, **8**(19), 2282–2292.

Mager, D. L., & Stoye, J. P. (2015). Mammalian Endogenous Retroviruses. In *Mobile DNA III*, Vol. 3, American Society of Microbiology, pp. 1079–1100.

Magiorkinis, G., Gifford, R. J., Katzourakis, A., De Ranter, J., & Belshaw, R. (2012). Env-less endogenous retroviruses are genomic superspreaders. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(19), 7385–7390.

Maksakova, I. A., Romanish, M. T., Gagnier, L., Dunn, C. A., Van De Lagemaat, L. N., & Mager, D. L. (2006). Retroviral elements and their hosts: Insertional mutagenesis in the mouse germ line. *PLoS Genetics*, **2**(1), 1–10.

Marikawa, Y., & Alarcón, V. B. (2009). Establishment of trophectoderm and inner cell mass lineages in the mouse embryo. *Molecular Reproduction and*

Development, **76**(11), 1019–1032.

Martens, J. H. a, O'Sullivan, R. J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., & Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *The EMBO Journal*, **24**(4), 800–812.

Martin, D. I. K., & Whitelaw, E. (1996). The vagaries of variegating transgenes. *BioEssays*, **18**(11), 919–923.

Matsui, T., Leung, D., Miyashita, H., Maksakova, I. a, Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M. C., & Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature*, **464**(7290), 927–931.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America*, **36**(6), 344–355.

McClintock, B. (1951). Chromosome organization and genic expression. *Cold Spring Harbor Symposia on Quantitative Biology*, **16**, 13–47.

McClintock, B. (1953). Induction of Instability at Selected Loci in Maize. *Genetics*, **38**(6), 579–99.

McClintock, B. (1956). Controlling elements and the gene. *Cold Spring Harbor Symposia on Quantitative Biology*, **21**(0), 197–216.

McClintock, B. (1958). The suppressor-mutator system of control of gene action in maize. *Carnegie Inst Wash Yearbook*, **57**, 415–429.

McLay, D. W., & Clarke, H. J. (2003). Remodelling the paternal chromatin at fertilization in mammals. *Reproduction (Cambridge, England)*, **125**(5), 625–33.

Meer, M. V, Podolskiy, D. I., Tyshkovskiy, A., & Gladyshev, V. N. (2018). A whole lifespan mouse multi-tissue DNA methylation clock. *eLife*, **7**, e40675.

Mi, S., Lee, X., Li, X., Veldman, G. M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.-Y., Edouard, P., Howes, S., Keith, J. C., & McCoy, J. M. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature*, **403**(6771), 785–789.

Michaud, E. J., van Vugt, M. J., Bultman, S. J., Sweet, H. O., Davisson, M. T., & Woychik, R. P. (1994). Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. *Genes & Development*, **8**(12), 1463–72.

Mieczkowski, P. A., Lemoine, F. J., & Petes, T. D. (2006). Recombination between

retrotransposons as a source of chromosome rearrangements in the yeast Saccharomyces cerevisiae. *DNA Repair*, **5**(9–10), 1010–1020.

Mikeska, T., & Craig, J. (2014). DNA Methylation Biomarkers: Cancer and Beyond. *Genes*, **5**(3), 821–864.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., … Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, **448**(7153), 553–560.

Miller, M. W., Duhl, D. M. J., Vrieling, H., Cordes, S. P., Ollmann, M. M., Winkes, B. M., & Barsh, G. S. (1993). Cloning of the mouse agouti gene predicts a secreted protein ubiquitously expressed in mice carrying the lethal yellow mutation. *Genes and Development*, **7**(3), 454–467.

Miltenberger, R. J., Mynatt, R. L., Wilkinson, J. E., & Woychik, R. P. (1997). The Role of the agouti Gene in the Yellow Obese Syndrome. *The Journal of Nutrition*, **127**(9), 1902–1907.

Miska, E. A., & Ferguson-Smith, A. C. (2016). Transgenerational inheritance: Models and mechanisms of non – DNA sequence – based inheritance. *Science*, **354**(6308), 778–782.

Miura, K., Agetsuma, M., Kitano, H., Yoshimura, A., Matsuoka, M., Jacobsen, S. E., & Ashikari, M. (2009). A metastable DWARF1 epigenetic mutant affecting plant stature in rice. *Proceedings of the National Academy of Sciences*, **106**(27), 11218–11223.

Molaro, A., Falciatori, I., Hodges, E., Aravin, A. A., Marran, K., Rafii, S., McCombie, W. R., Smith, A. D., & Hannon, G. J. (2014). Two waves of de novo methylation during mouse germ cell development. *Genes & Development*, **28**(14), 1544–1549.

Monk, M., Boubelik, M., & Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development*, **99**(3).

Montrose, L., Faulk, C., Francis, J., & Dolinoy, D. C. (2017). Perinatal lead (Pb) exposure results in sex and tissue-dependent adult DNA methylation alterations in murine IAP transposons. *Environmental and Molecular Mutagenesis*, **58**(8), 540–550.

Morgan, A. P., Fu, C. P., Kao, C. Y., … Villena, F. P. M. de. (2016). The mouse universal genotyping array: From substrains to subspecies. *G3: Genes, Genomes, Genetics*, **6**(2), 263–279.

Morgan, H. D., Jin, X. L., Li, A., Whitelaw, E., & O'Neill, C. (2008). The culture of

zygotes to the blastocyst stage changes the postnatal expression of an epigentically labile allele, agouti viable yellow, in mice. *Biology of Reproduction*, **79**(4), 618–23.

Morgan, H. D., Sutherland, H. G., Martin, D. I., & Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. *Nature Genetics*, **23**(3), 314–318.

Neier, K., Cheatham, D., Bedrosian, L. D., & Dolinoy, D. C. (2019). Perinatal exposures to phthalates and phthalate mixtures result in sex-specific effects on body weight, organ weights and intracisternal A-particle (IAP) DNA methylation in weanling mice. *Journal of Developmental Origins of Health and Disease*, **10**(2), 176–187.

Nellåker, C., Keane, T. M., Yalcin, B., Wong, K., Agam, A., Belgard, T. G., Flint, J., Adams, D. J., Frankel, W. N., & Ponting, C. P. (2012). The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biology*, **13**(6), R45.

Nelson, V. R., Spezio, S. H., & Nadeau, J. H. (2010). Transgenerational genetic effects of the paternal y chromosome on daughters phenotypes. *Epigenomics*, **2**(4), 513–521.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., & Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, **543**(7643), 72–77.

Ng, S.-F., Lin, R. C. Y., Laybutt, D. R., Barres, R., Owens, J. A., & Morris, M. J. (2010). Chronic high-fat diet in fathers programs β-cell dysfunction in female rat offspring. *Nature*, **467**(7318), 963–966.

Oey, H., Isbel, L., Hickey, P., Ebaid, B., & Whitelaw, E. (2015). Genetic and epigenetic variation among inbred mouse littermates: identification of inter-individual differentially methylated regions. *Epigenetics & Chromatin*, **8**(1), 54.

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell*, **99**(3), 247–257.

Okano, M., Xie, S., & Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genetics*, **19**(3), 219–220.

Onuchic, V., Lurie, E., Carrero, I., … Milosavljevic, A. (2018). Allele-specific epigenome maps reveal sequence-dependent stochastic switching at regulatory loci. *Science*, **361**(6409), eaar3146.

Ooi, S. K. T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C. D., Cheng, X., & Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, **448**(7154), 714–717.

Opsahl, M. L., McClenaghan, M., Springbett, A., Reid, S., Lathe, R., Colman, A., & Whitelaw, C. B. A. (2002). Multiple effects of genetic background on variegated transgene expression in mice. *Genetics*, **160**(3), 1107–1112.

Orgel, L. E., & Crick, F. H. C. (1980). Selfish DNA: the ultimate parasite. *Nature*, **284**(5757), 604–607.

Ounpraseuth, S., Rafferty, T. M., McDonald-Phillips, R. E., Gammill, W. M., Siegel, E. R., Wheeler, K. L., Nilsson, E. A., & Cooney, C. A. (2009). A Method to Quantify Mouse Coat-Color Proportions. *PLoS ONE*, **4**(4), e5414.

Padmanabhan, N., Jia, D., Geary-Joo, C., Wu, X., Ferguson-Smith, A. C., Fung, E., Bieda, M. C., Snyder, F. F., Gravel, R. A., Cross, J. C., & Watson, E. D. (2013). Mutation in Folate Metabolism Causes Epigenetic Instability and Transgenerational Effects on Development. *Cell*, **155**(1), 81–93.

Padmanabhan, N., Menelaou, K., Gao, J., Anderson, A., Blake, G. E. T., Li, T., Daw, B. N., & Watson, E. D. (2018). Abnormal folate metabolism causes age-, sex- and parent-of-origin-specific haematological defects in mice. *The Journal of Physiology*, **596**(18), 4341–4360.

Padmanabhan, N., Rakoczy, J., Kondratowicz, M., Menelaou, K., Blake, G. E. T., & Watson, E. D. (2017). Multigenerational analysis of sex-specific phenotypic differences at midgestation caused by abnormal folate metabolism. *Environmental Epigenetics*, **3**(4), dvx014.

Peaston, A. E., Evsikov, A. V., Graber, J. H., de Vries, W. N., Holbrook, A. E., Solter, D., & Knowles, B. B. (2004). Retrotransposons Regulate Host Genes in Mouse Oocytes and Preimplantation Embryos. *Developmental Cell*, **7**(4), 597–606.

Perk, K., & Dahlberg, J. E. (1974). Murine intracisternal A type particles fail to separate from the membrane of the endoplasmic reticulum. *Journal of Virology*, **14**(5), 1304–6.

Petkovich, D. A., Podolskiy, D. I., Lobanov, A. V, Lee, S., Miller, R. A., & Gladyshev, V. N. (2017). Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. *Cell Metabolism*, **25**(4), 954–960.

Pezic, D., Manakov, S. A., Sachidanandam, R., & Aravin, A. A. (2014). piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3

mark in germ cells. *Genes & Development*, **28**(13), 1410–28.

Phillips, J. E., & Corces, V. G. (2009). CTCF: Master Weaver of the Genome. *Cell*, **137**(7), 1194–1211.

Preis, J. I., Downes, M., Oates, N. A., Rasko, J. E. J., & Whitelaw, E. (2003). Sensitive flow cytometric analysis reveals a novel type of parent-of-origin effect in the mouse genome. *Current Biology*, **13**(11), 955–959.

Qin, C., Wang, Z., Shang, J., Bekkari, K., Liu, R., Pacchione, S., McNulty, K. A., Ng, A., Barnum, J. E., & Storer, R. D. (2010). Intracisternal a particle genes: Distribution in the mouse genome, active subtypes, and potential roles as species-specific mediators of susceptibility to cancer. *Molecular Carcinogenesis*, **49**(1), 54–67.

Quadrana, L., & Colot, V. (2016). Plant Transgenerational Epigenetics. *Annual Review of Genetics*, **50**(1), 467–491.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**(6), 841–842.

Radford, E. J. (2018). Exploring the extent and scope of epigenetic inheritance. *Nature Reviews Endocrinology*, **14**(6), 345–355.

Radford, E. J., Ito, M., Shi, H., Corish, J. A., Yamazawa, K., Isganaitis, E., Seisenberger, S., Hore, T. A., Reik, W., Erkek, S., Peters, A. H. F. M., Patti, M.-E., & Ferguson-Smith, A. C. (2014). In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. *Science*, **345**(6198), 1255903.

Raj, A., & van Oudenaarden, A. (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*, **135**(2), 216–226.

Rakyan, V. K., Blewitt, M. E., Druker, R., Preis, J. I., & Whitelaw, E. (2002). Metastable epialleles in mammals. *Trends in Genetics*, **18**(7), 348–351.

Rakyan, V. K., Chong, S., Champ, M. E., Cuthbert, P. C., Morgan, H. D., Luu, K. V. K., & Whitelaw, E. (2003). Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(5), 2538–2543.

Ramos, S. B. V, Stumpo, D. J., Kennington, E. A., Phillips, R. S., Bock, C. B., Ribeiro-Neto, F., & Blackshear, P. J. (2004). The CCCH tandem zinc-finger protein Zfp36l2 is crucial for female fertility and early embryonic development. *Development*, **131**(19), 4883–93.

Rangasamy, D. (2013). Distinctive patterns of epigenetic marks are associated with promoter regions of mouse LINE-1 and LTR retrotransposons. *Mobile DNA*, **4**(1), 27.

Ratnam, S., Bozek, G., Martin, T., Gallagher, S. J., Payne, C. J., & Storb, U. (2017). Ssm1b expression and function in germ cells of adult mice and in early embryos. *Molecular Reproduction and Development*, **84**(7), 596–613.

Ratnam, S., Engler, P., Bozek, G., Mao, L., Podlutsky, A., Austad, S., Martin, T., & Storb, U. (2014). Identification of Ssm1b, a novel modifier of DNA methylation, and its expression during mouse embryogenesis. *Development*, **141**(10), 2024–2034.

Rebollo, R., Romanish, M. T., & Mager, D. L. (2012). Transposable Elements: An Abundant and Natural Source of Regulatory Sequences for Host Genes. *Annual Review of Genetics*, **46**, 21–42.

Reed, S. C. (1937). The Inheritance and Expression of Fused, a New Mutation in the House Mouse. *Genetics*, **22**(1), 1–13.

Reik, W., Collick, A., Norris, M. L., Barton, S. C., & Surani, M. A. (1987). Genomic imprinting determines methylation of parental alleles in transgenic mice. *Nature*, **328**(6127), 248–51.

Reiss, D., & Mager, D. L. (2007). Stochastic epigenetic silencing of retrotransposons: Does stability come with age? *Gene*, **390**(1–2), 130–135.

Reuss, F. U., & Schaller, H. C. (1991). cDNA sequence and genomic characterization of intracisternal A-particle-related retroviral elements containing an envelope gene. *Journal of Virology*, **65**(11), 5702–9.

Ribet, D., Harper, F., Dupressoir, A., Dewannieux, M., Pierron, G., & Heidmann, T. (2008). An infectious progenitor for the murine IAP retrotransposon: Emergence of an intracellular genetic parasite from an ancient retrovirus. *Genome Research*, **18**(4), 597–609.

Riggs, A. D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenetic and Genome Research*, **14**(1), 9–25.

Roach, W. G., Chavez, J. A., Mîinea, C. P., & Lienhard, G. E. (2007). Substrate specificity and effect on GLUT4 translocation of the Rab GTPase-activating protein Tbc1d1. *The Biochemical Journal*, **403**(2), 353–8.

Robbez-Masson, L., Tie, C. H. C., Conde, L., Tunbak, H., Husovsky, C., Tchasovnikarova, I. A., Timms, R. T., Herrero, J., Lehner, P. J., & Rowe, H. M. (2018). The hush complex cooperates with trim28 to repress young retrotransposons and new genes. *Genome Research*, **28**(6), 836–845.

Rojas-Ríos, P., & Simonelig, M. (2018). piRNAs and PIWI proteins: Regulators of gene expression in development and stem cells. *Development*, **145**(17), dev161786.

Roovers, E. F., Rosenkranz, D., Mahdipour, M., Han, C.-T., He, N., Chuva de Sousa Lopes, S. M., van der Westerlaken, L. A. J., Zischler, H., Butter, F., Roelen, B. A. J., & Ketting, R. F. (2015). Piwi Proteins and piRNAs in Mammalian Oocytes and Early Embryos. *Cell Reports*, **10**(12), 2069–2082.

Rosenfeld, C. S., Sieli, P. T., Warzak, D. A., Ellersieck, M. R., Pennington, K. A., & Roberts, R. M. (2013). Maternal exposure to bisphenol A and genistein has minimal effect on A(vy)/a offspring coat color but favors birth of agouti over nonagouti mice. *Proceedings of the National Academy of Sciences of the United States of America*, **110**(2), 537–42.

Rowe, H. M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V, Layard-Liesching, H., Verp, S., Marquis, J., Spitz, F., Constam, D. B., & Trono, D. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*, **463**(7278), 237–240.

Rowe, H. M., & Trono, D. (2011). Dynamic control of endogenous retroviruses during development. *Virology*, **411**(2), 273–287.

Ruvinsky, A. O., & Agulnik, A. I. (1990). Gametic imprinting and the manifestation of the fused gene in the house mouse. *Developmental Genetics*, **11**(4), 263–269.

Samuelsson, A.-M., Matthews, P. A., Argenton, M., Christie, M. R., McConnell, J. M., Jansen, E. H. J. M., Piersma, A. H., Ozanne, S. E., Twinn, D. F., Remacle, C., Rowlerson, A., Poston, L., & Taylor, P. D. (2008). Diet-Induced Obesity in Female Mice Leads to Offspring Hyperphagia, Adiposity, Hypertension, and Insulin Resistance. *Hypertension*, **51**(2), 383–392.

Sandovici, I., Kassovska-Bratinova, S., Loredo-Osti, J. C., Leppert, M., Suarez, A., Stewart, R., Bautista, F. D., Schiraldi, M., & Sapienza, C. (2005). Interindividual variability and parent of origin DNA methylation differences at specific human Alu elements. *Human Molecular Genetics*, **14**(15), 2135–2143.

Sanford, J. P., Clark, H. J., Chapman, V. M., & Rossant, J. (1987). Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. *Genes & Development*, **1**(10), 1039–46.

Santos, F., Hendrich, B., Reik, W., & Dean, W. (2002). Dynamic Reprogramming of DNA Methylation in the Early Mouse Embryo. *Developmental Biology*,

241(1), 172–182.

Sapienza, C., Paquette, J., Tran, T. H., & Peterson, A. (1989). Epigenetic and genetic factors affect transgene methylation imprinting. *Development*, **107**(1), 165–168.

Sapienza, C., Peterson, A. C., Rossant, J., & Balling, R. (1987). Degree of methylation of transgenes is dependent on gamete of origin. *Nature*, **328**(6127), 251–254.

Sasaki, H., & Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: Reprogramming and beyond. *Nature Reviews Genetics*, **9**(2), 129–140.

Schnable, P. S., Ware, D., Fulton, R. S., … Wilson, R. K. (2009). The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science*, **326**(5956), 1112–1115.

Schotta, G., Ebert, A., Dorn, R., & Reuter, G. (2003). Position-effect variegation and the genetic dissection of chromatin regulation in Drosophila. *Seminars in Cell and Developmental Biology*, **14**(1), 67–75.

Schroeder, D. I., Jayashankar, K., Douglas, K. C., Thirkill, T. L., York, D., Dickinson, P. J., Williams, L. E., Samollow, P. B., Ross, P. J., Bannasch, D. L., Douglas, G. C., & LaSalle, J. M. (2015). Early Developmental and Evolutionary Origins of Gene Body DNA Methylation Patterns in Mammalian Placentas. *PLoS Genetics*, **11**(8), 1–20.

Schübeler, D. (2015). Function and information content of DNA methylation. *Nature*, **517**(7534), 321–326.

Seah, M. K. Y., Wang, Y., Goy, P.-A., Loh, H. M., Peh, W. J., Low, D. H. P., Han, B. Y., Wong, E., Leong, E. L., Wolf, G., Mzoughi, S., Wollmann, H., Macfarlan, T. S., Guccione, E., & Messerschmidt, D. M. (2019). The KRAB-zinc-finger protein ZFP708 mediates epigenetic repression at RMER19B retrotransposons. *Development*, **146**(19), dev.170266.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., & Reik, W. (2012). The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Molecular Cell*, **48**(6), 849–862.

Seisenberger, S., Peat, J. R., Hore, T. A., Santos, F., Dean, W., & Reik, W. (2013). Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **368**(1609), 20110330.

Seperack, P. K., Strobel, M. C., Corrow, D. J., Jenkins, N. A., & Copeland, N. G. (2006). Somatic and germ-line reverse mutation rates of the retrovirus-induced dilute coat-color mutation of DBA mice. *Proceedings of the National Academy of Sciences*, **85**(1), 189–192.

Shane, B., & Stokstad, E. L. R. (1985). Vitamin B12-Folate Interrelationships. *Annual Review of Nutrition*, **5**(1), 115–141.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T. A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., Tajima, S., Mitsuya, K., Okano, M., & Koseki, H. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, **450**(7171), 908–912.

Sharif, J., Shinkai, Y., & Koseki, H. (2013). Is there a role for endogenous retroviruses to mediate long-term adaptive phenotypic response upon environmental inputs? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **368**(1609), 20110340.

Shen, L., Inoue, A., He, J., Liu, Y., Lu, F., & Zhang, Y. (2014). Tet3 and DNA Replication Mediate Demethylation of Both the Maternal and Paternal Genomes in Mouse Zygotes. *Cell Stem Cell*, **15**(4), 459–471.

Shimosuga, K. I., Fukuda, K., Sasaki, H., & Ichiyanagi, K. (2017). Locus-specific hypomethylation of the mouse IAP retrotransposon is associated with transcription factor-binding sites. *Mobile DNA*, **8**(1), 1–12.

Shorter, K. R., Anderson, V., Cakora, P., Owen, A., Lo, K., Crossland, J., South, A. C. H., Felder, M. R., & Vrana, P. B. (2014). Pleiotropic effects of a methyl donor diet in a novel animal model. *PLoS ONE*, **9**(8), e104942.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., & Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*, **479**(7371), 74–79.

Silver, M. J., Kessler, N. J., Hennig, B. J., Dominguez-Salas, P., Laritsky, E., Baker, M. S., Coarfa, C., Hernandez-Vargas, H., Castelino, J. M., Routledge, M. N., Gong, Y. Y., Herceg, Z., Lee, Y. S., Lee, K., Moore, S. E., Fulford, A. J., Prentice, A. M., & Waterland, R. A. (2015). Independent genomewide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment. *Genome Biology*, **16**(1), 1–14.

Slotkin, R. K., & Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics*, **8**(4), 272–285.

Smallwood, S. A., & Kelsey, G. (2012). De novo DNA methylation: A germ cell perspective. *Trends in Genetics*, **28**(1), 33–42.

Smit, A. F. A., Hubley, R., & Green, P. (2015). RepeatMasker Open-4.0. Retrieved April 25, 2019, from http://www.repeatmasker.org/

Sorolla, A., Tallack, M. R., Oey, H., Harten, S. K., Daxinger, L. C., Magor, G. W., Combes, A. N., Ilsley, M., Whitelaw, E., & Perkins, A. C. (2015). Identification of novel hypomorphic and null mutations in Klf1 derived from a genetic screen for modifiers of α-globin transgene variegation. *Genomics*, **105**(2), 116–122.

Stephens, A. S., Stephens, S. R., & Morrison, N. A. (2011). Internal control genes for quantitative RT-PCR expression analysis in mouse osteoblasts , osteoclasts and macrophages. *BMC Research Notes*, **4**(1), 410.

Stoye, J. P., Fenner, S., Greenoak, G. E., Moran, C., & Coffin, J. M. (1988). Role of endogenous retroviruses as mutagens: The hairless mutation of mice. *Cell*, **54**(3), 383–391.

Strogantsev, R., Krueger, F., Yamazawa, K., Shi, H., Gould, P., Goldman-Roberts, M., McEwen, K., Sun, B., Pedersen, R., & Ferguson-Smith, A. C. (2015). Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. *Genome Biology*, **16**(1), 112.

Stubbs, T. M., Bonder, M. J., Stark, A.-K., Krueger, F., von Meyenn, F., Stegle, O., & Reik, W. (2017). Multi-tissue DNA methylation age predictor in mouse. *Genome Biology*, **18**(1), 68.

Styles, E., & Brink, R. (1966). The Metastable Nature of Paramutable R Alleles in Maize. I. Heritable Enhancement in Level of Standard Rr Action. *Genetics*, **54**(1077), 433–439.

Subramanian, R. P., Wildschutte, J. H., Russo, C., & Coffin, J. M. (2011). Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology*, **8**(1), 90.

Sun, Q., Cornelis, M. C., Townsend, M. K., Tobias, D. K., Eliassen, A. H., Franke, A. A., Hauser, R., & Hu, F. B. (2014). Association of Urinary Concentrations of Bisphenol A and Phthalate Metabolites with Risk of Type 2 Diabetes: A Prospective Investigation in the Nurses' Health Study (NHS) and NHSII Cohorts. *Environmental Health Perspectives*, **122**(6), 616–623.

Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M. P., & Wang, T. (2014). Widespread contribution of transposable elements to the

innovation of gene regulatory networks. *Genome Research*, **24**(12), 1963–1976.

Susiarjo, M., Sasson, I., Mesaros, C., & Bartolomei, M. S. (2013). Bisphenol A Exposure Disrupts Genomic Imprinting in the Mouse. *PLoS Genetics*, **9**(4), e1003401.

Susiarjo, M., Xin, F., Bansal, A., Stefaniak, M., Li, C., Simmons, R. A., & Bartolomei, M. S. (2015). Bisphenol A Exposure Disrupts Metabolic Health Across Multiple Generations in the Mouse. *Endocrinology*, **156**(6), 2049–2058.

Sutherland, H. G. E., Kearns, M., Morgan, H. D., Headley, A. P., Morris, C., Martin, D. I. K., & Whitelaw, E. (2000). Reactivation of heritably silenced gene expression in mice. *Mammalian Genome*, **11**(5), 347–355.

Svoboda, P., Stein, P., Anger, M., Bernstein, E., Hannon, G. J., & Schultz, R. M. (2004). RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Developmental Biology*, **269**(1), 276–285.

Swain, J. L., Stewart, T. A., & Leder, P. (1987). Parental legacy determines methylation and expression of an autosomal transgene: A molecular mechanism for parental imprinting. *Cell*, **50**(5), 719–727.

Takahashi, N., Coluccio, A., Thorball, C. W., Planet, E., Shi, H., Offner, S., Turelli, P., Imbeault, M., Ferguson-Smith, A. C., & Trono, D. (2019). ZNF445 is a primary regulator of genomic imprinting. *Genes & Development*, **33**(1–2), 49–54.

Tam, O. H., Aravin, A. A., Stein, P., Girard, A., Murchison, E. P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R. M., & Hannon, G. J. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature*, **453**(7194), 534–538.

Tan, X., Xu, X., Elkenani, M., Smorag, L., Zechner, U., Nolte, J., Engel, W., & Pantakani, D. V. K. (2013). Zfp819, a novel KRAB-zinc finger protein, interacts with KAP1 and functions in genomic integrity maintenance of mouse embryonic stem cells. *Stem Cell Research*, **11**(3), 1045–1059.

Tchasovnikarova, I. A., Timms, R. T., Douse, C. H., Roberts, R. C., Dougan, G., Kingston, R. E., Modis, Y., & Lehner, P. J. (2017). Hyperactivation of HUSH complex function by Charcot–Marie–Tooth disease mutation in MORC2. *Nature Genetics*, **49**(7), 1035–1044.

Tchasovnikarova, I. A., Timms, R. T., Matheson, N. J., Wals, K., Antrobus, R., Göttgens, B., Dougan, G., Dawson, M. A., & Lehner, P. J. (2015). Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science*, **348**(6242), 1481–1485.

Thomas, J. H., & Schneider, S. (2011). Coevolution of retroelements and tandem zinc finger genes. *Genome Research*, **21**(11), 1800–1812.

Thomas, J., Perron, H., & Feschotte, C. (2018). Variation in proviral content among human genomes mediated by LTR recombination 06 Biological Sciences 0604 Genetics. *Mobile DNA*, **9**(1), 36.

Thompson, P. J., Macfarlan, T. S., & Lorincz, M. C. (2016). Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. *Molecular Cell*, **62**(5), 766–776.

Timmons, J. A., Atherton, P. J., Larsson, O., Sood, S., Blokhin, I. O., Brogan, R. J., Volmar, C.-H., Josse, A. R., Slentz, C., Wahlestedt, C., Phillips, S. M., Phillips, B. E., Gallagher, I. J., & Kraus, W. E. (2018). A coding and non-coding transcriptomic perspective on the genomics of human metabolic disease. *Nucleic Acids Research*, **46**(15), 7772–7792.

Ting, C. N., Rosenberg, M. P., Snow, C. M., Samuelson, L. C., & Meisler, M. H. (1992). Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes & Development*, **6**(8), 1457–65.

Tost, J., & Gut, I. G. (2007). DNA methylation analysis by pyrosequencing. *Nature Protocols*, **2**(9), 2265–2275.

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J. G., Baylin, S. B., & Issa, J.-P. J. (1999). CpG island methylator phenotype in colorectal cancer. *Proceedings of the National Academy of Sciences*, **96**(15), 8681–8686.

Tremblay, K. D., Duran, K. L., & Bartolomei, M. S. (1997). A 5′ 2-kilobase-pair region of the imprinted mouse H19 gene exhibits exclusive paternal methylation throughout development. *Molecular and Cellular Biology*, **17**(8), 4322–4329.

Turelli, P., Castro-Diaz, N., Marzetta, F., Kapopoulou, A., Raclot, C., Duc, J., Tieng, V., Quenneville, S., & Trono, D. (2014). Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Research*, **24**(8), 1260–70.

Van Baak, T. E., Coarfa, C., Dugué, P.-A., … Waterland, R. A. (2018). Epigenetic supersimilarity of monozygotic twin pairs. *Genome Biology*, **19**(1), 2.

Varmus, H. E. (1982). Form and function of retroviral proviruses. *Science*, **216**(4548), 812–20.

Vasicek, T. J., Zeng, L., Guan, X. J., Zhang, T., Costantini, F., & Tilghman, S. M. (1997). Two dominant mutations in the mouse fused gene are the result of

transposon insertions. *Genetics*, **147**(2), 777–86.

Vincent, J. J., Huang, Y., Chen, P.-Y., Feng, S., Calvopiña, J. H., Nee, K., Lee, S. A., Le, T., Yoon, A. J., Faull, K., Fan, G., Rao, A., Jacobsen, S. E., Pellegrini, M., & Clark, A. T. (2013). Stage-Specific Roles for Tet1 and Tet2 in DNA Demethylation in Primordial Germ Cells. *Cell Stem Cell*, **12**(4), 470–478.

Walsh, C. P., Chaillet, J. R., & Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genetics*, **20**(2), 116–7.

Walter, M., Teissandier, A., Pérez-Palacios, R., & Bourc'his, D. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife*, **5**, e11418.

Wang, H., Maurano, M. T., Qu, H., Varley, K. E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., Thurman, R. E., Kaul, R., Myers, R. M., & Stamatoyannopoulos, J. A. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Research*, **22**(9), 1680–1688.

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., Li, W., Zhou, Q., Aluru, N., Tang, F., He, C., Huang, X., & Liu, J. (2014). Programming and Inheritance of Parental DNA Methylomes in Mammals. *Cell*, **157**(4), 979–991.

Wang, R. Y.-H., Kuo, K. C., Gehrke, C. W., Huang, L.-H., & Ehrlich, M. (1982). Heat- and alkali-induced deamination of 5-methylcytosine and cytosine residues in DNA. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, **697**(3), 371–377.

Wang, T., Tsui, B., Kreisberg, J. F., Robertson, N. A., Gross, A. M., Yu, M. K., Carter, H., Brown-Borg, H. M., Adams, P. D., & Ideker, T. (2017). Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. *Genome Biology*, **18**(1), 57.

Wang, Z., McSwiggin, H., Newkirk, S. J., Wang, Y., Oliver, D., Tang, C., Lee, S., Wang, S., Yuan, S., Zheng, H., Ye, P., An, W., & Yan, W. (2019). Insertion of a chimeric retrotransposon sequence in mouse Axin1 locus causes metastable kinky tail phenotype. *Mobile DNA*, **10**(1), 17.

Warzak, D. A., Johnson, S. A., Ellersieck, M. R., Roberts, R. M., Zhang, X., Ho, S. M., & Rosenfeld, C. S. (2015). Effects of post-weaning diet on metabolic parameters and DNA methylation status of the cryptic promoter in the A^{vy} allele of viable yellow mice. *Journal of Nutritional Biochemistry*, **26**(6), 667–674.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., Surani, M. A., Sakaki, Y., & Sasaki, H. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, **453**(7194), 539–543.

Waterland, R. A. (2006). Assessing the effects of high methionine intake on DNA methylation. *The Journal of Nutrition*, **136**(6), 1706S-1710S.

Waterland, R. A., Dolinoy, D. C., Lin, J. R., Smith, C. A., Shi, X., & Tahiliani, K. G. (2006). Maternal Methyl Supplements Increase Offspring DNA Methylation at Axin Fused. *Genesis*, **44**(9), 401–406.

Waterland, R. A., & Jirtle, R. L. (2003). Transposable Elements : Targets for Early Nutritional Effects on Epigenetic Gene Regulation. *Molecular and Cellular Biology*, **23**(15), 5293–5300.

Waterland, R. A., Kellermayer, R., Laritsky, E., Rayco-Solon, P., Harris, R. A., Travisano, M., Zhang, W., Torskaya, M. S., Zhang, J., Shen, L., Manary, M. J., & Prentice, A. M. (2010). Season of Conception in Rural Gambia Affects DNA Methylation at Putative Human Metastable Epialleles. *PLoS Genetics*, **6**(12), e1001252.

Waterland, R. A., Travisano, M., & Tahiliani, K. G. (2007). Diet-induced hypermethylation at agouti viable yellow is not inherited transgenerationally through the female. *The FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, **21**(12), 3380–3385.

Weidner, C. I., Lin, Q., Koch, C. M., Eisele, L., Beier, F., Ziegler, P., Bauerschlag, D. O., Jöckel, K.-H., Erbel, R., Mühleisen, T. W., Zenke, M., Brümmendorf, T. H., & Wagner, W. (2014). Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biology*, **15**(2), R24.

Weinhouse, C., Anderson, O. S., Jones, T. R., Kim, J., Liberman, S. a, Nahar, M. S., Rozek, L. S., Jirtle, R. L., & Dolinoy, D. C. (2011). An expression microarray approach for the identification of metastable epialleles in the mouse genome. *Epigenetics : Official Journal of the DNA Methylation Society*, **6**(9), 1105–13.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, **8**(12), 973–982.

Wolf, D., & Goff, S. P. (2009). Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature*, **458**(7242), 1201–1204.

Wolf, G., Yang, P., Füchtbauer, A. C., Füchtbauer, E. M., Silva, A. M., Park, C., Wu, W., Nielsen, A. L., Pedersen, F. S., & Macfarlan, T. S. (2015). The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes and Development*, **29**(5), 538–554.

Wolff, G. L. (1971). Genetic modification of homeostatic regulation in the mouse. *The American Naturalist*, **105**(943), 241–252.

Wolff, G. L. (1978). Influence of maternal phenotype on metabolic differentiation of agouti locus mutants in the mouse. *Genetics*, **88**(3), 529–39.

Wolff, G. L., Kodell, R. L., Moore, S. R., & Cooney, C. A. (1998). Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *The FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, **12**(11), 949–957.

Yamada, K., Gravel, R. A., Toraya, T., & Matthews, R. G. (2006). Human methionine synthase reductase is a molecular chaperone for human methionine synthase. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(25), 9476–81.

Yang, X., Lay, F., Han, H., & Jones, P. A. (2010). Targeting DNA methylation for epigenetic therapy. *Trends in Pharmacological Sciences*, **31**(11), 536–546.

Yen, T. T., Gill, A. M., Frigeri, L. G., Barsh, G. S., & Wolff, G. L. (1994). Obesity, diabetes, and neoplasia in yellow A(vy)/- mice: ectopic expression of the agouti gene. *The FASEB Journal*, **8**(8), 479–488.

Yoder, J. A., Walsh, C. P., & Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics*, **13**(8), 335–340.

Youngson, N. A., Vickaryous, N., Van Der Horst, A., Epp, T., Harten, S., Fleming, J. S., Khanna, K. K., De Kretser, D. M., & Whitelaw, E. (2011). A missense mutation in the transcription factor Foxo3a causes teratomas and oocyte abnormalities in mice. *Mammalian Genome*, **22**(3–4), 235–248.

Zemach, A., Mcdaniel, I., Silva, P., & Zilberman, D. (2010). Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. *Science*, **328**(5980), 916–919.

Zemach, A., & Zilberman, D. (2010). Evolution of Eukaryotic DNA Methylation and the Pursuit of Safer Sex. *Current Biology*, **20**(17), R780–R785.

Zeng, L., Fagotto, F., Zhang, T., Hsu, W., Vasicek, T. J., Perry, W. L., Lee, J. J., Tilghman, S. M., Gumbiner, B. M., & Costantini, F. (1997). The mouse Fused locus encodes axin, an inhibitor of the Wnt signaling pathway that regulates embryonic axis formation. *Cell*, **90**(1), 181–192.

Zeng, Y., Wang, W., Ma, J., Wang, X., Guo, M., & Li, W. (2012). Knockdown of ZNF268, which is transcriptionally downregulated by GATA-1, promotes proliferation of K562 cells. *PloS One*, **7**(1), e29518.

Zhang, Y., Maksakova, I. A., Gagnier, L., Van De Lagemaat, L. N., & Mager, D. L. (2008). Genome-wide assessments reveal extremely high levels of polymorphism of two active families of mouse endogenous retroviral elements. *PLoS Genetics*, **4**(2), e1000007.

Zhao, Y., & Garcia, B. A. (2015). Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harbor Perspectives in Biology*, **7**(9), a025064.

Appendix A

Complementary information and data

Table A.1: VM-IAP genomic coordinates.

Locus	GRCm38/mm10 coordinates
VM-IAP$_{Mbnl1}$	chr3:60,489,874-60,495,084
VM-IAP$_{Tfpi}$	chr2:84,505,211-84,510,421
VM-IAP$_{Bmf}$	chr2:118,554,767-118,558,375
VM-IAP$_{Rnf157}$	chr11:116,389,406-116,394,709
VM-IAP$_{Eps8l1}$	chr7:4,455,443-4,460,702
VM-IAP$_{Diap3}$	chr14:86,398,119-86,398,596
VM-IAP$_{Fam78b}$	chr1:166,940,023-166,940,493
VM-IAP$_{Marveld2}$	chr13:100,588,953-100,594,225
VM-IAP$_{Rab6b}$	chr9:103,108,603-103,109,112
VM-IAP$_{Sema6d}$	chr2:123,754,522-123,754,992
VM-IAP$_{Gm13849}$	chr6:31,799,371-31,804,637
VM-IAP$_{Slc15a2}$	chr16:36,765,056-36,770,352
VM-IAP$_{Pink1}$	chr4:138,327,386-138,327,850
VM-IAP$_{Ect2l}$	chr10:18,139,121-18,139,597

Table A.2: IAP genomic coordinates.

Locus	GRCm38/mm10 coordinates
IAP_{Cdc73}	chr1:142584892-142590150
IAP_{H2T24}	chr17:36007144-36014240
IAP_{Diap3}	chr14:85889096-85896207
IAP_{Dst}	chr1:34347457-34351369
IAP_{Cdk15}	chr1:59345147-59350179
IAP_{Ell2}	chr13:75677018-75682366
$IAP_{Slc24a2}$	chr2:145118404-145123718
IAP_{Zak}	chr2:72270362-72275599

Table A.3: Normality tests for VM-IAP methylation frequency distributions.

	VM-IAP					
Test	Tfpi	Mbnl1	Rnf157	Gm13849	Marveld2	Bmf
Anderson-Darling test						
Test statistic (A^2)	4.100	2.255	3.165	3.232	2.007	3.231
P-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Passed normality test?[1]	No	No	No	No	No	No
D'Agostino & Pearson test						
Test statistic (K^2)	24.520	32.870	14.660	13.720	7.380	26.580
P-value	<0.0001	<0.0001	0.0007	0.0011	0.025	<0.0001
Passed normality test?[1]	No	No	No	No	No	No
Shapiro-Wilk test						
Test statistic (W)	0.956	0.956	0.950	0.949	0.969	0.951
P-value	<0.0001	<0.0001	<0.0001	<0.0001	0.0004	<0.0001
Passed normality test?[1]	No	No	No	No	No	No
Kolmogorov-Smirnov test						
Test statistic (K^2 distance)	0.097	0.072	0.091	0.085	0.128	0.083
P-value	<0.0001	0.0019	0.0002	0.0001	<0.0001	<0.0001
Passed normality test?[1]	No	No	No	No	No	No
Sample size (individuals)	300	265	213	262	183	278

[1]alpha = 0.05

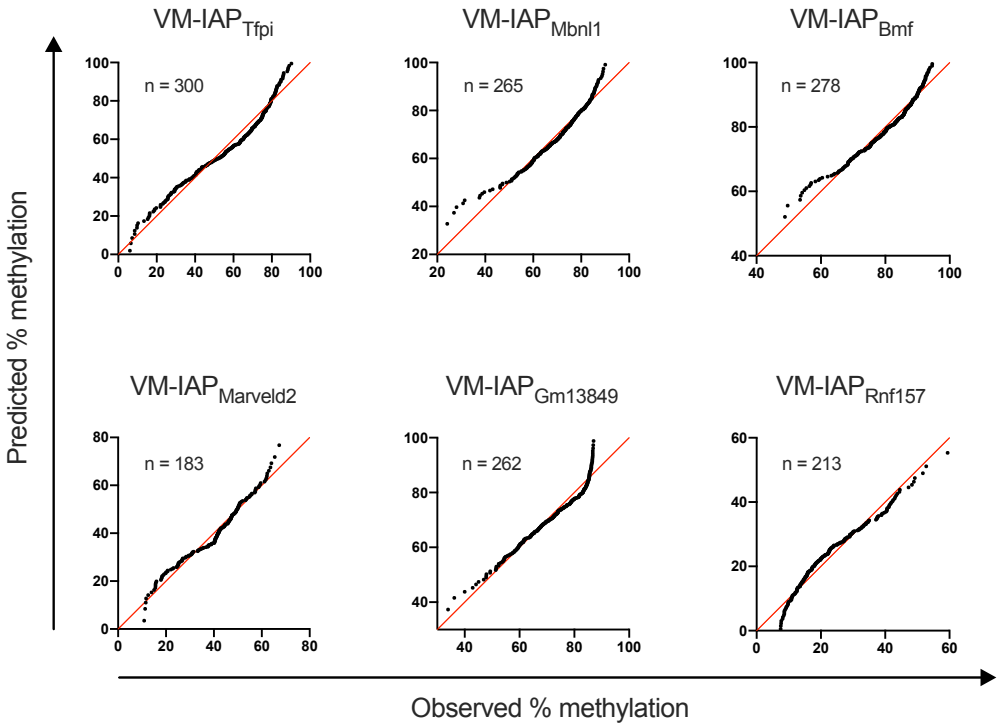


Figure A.1: Quantile-quantile plots for VM-IAP methylation frequency distributions. Y axes show the predicted percent methylation values assuming sampling from a Gaussian distribution. X axes show the empirically observed percent methylation values. Methylation values were averaged across CpGs at each. Sample sizes are shown on the graphs. The red line represents the line of identity.

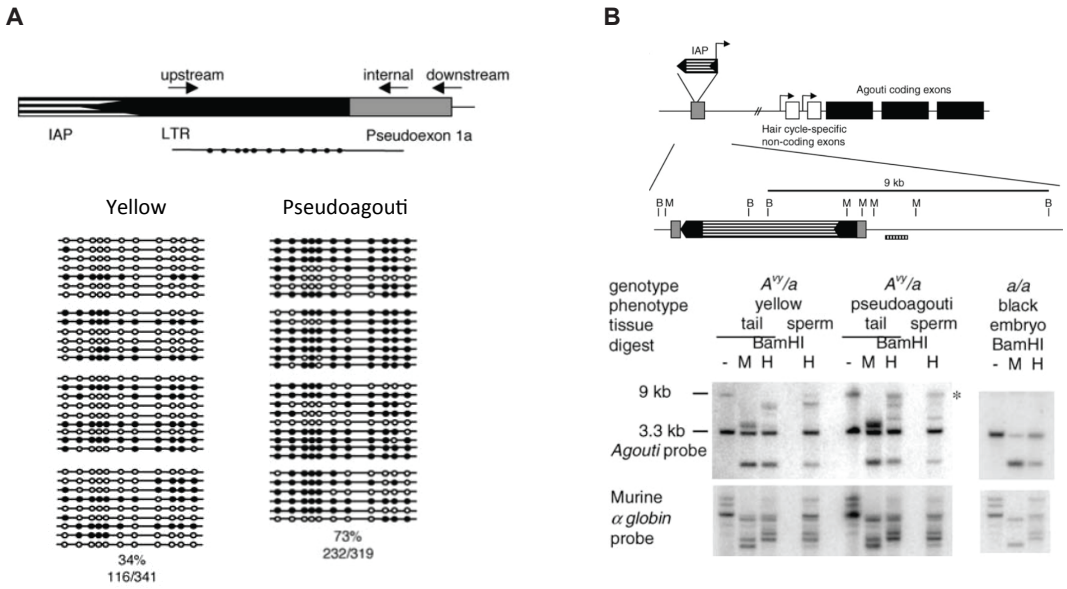


Figure A.2: Previously published data on A^{vy} methylation in mature sperm. A. Clonal bisulphite sequencing of yellow and pseudoagouti A^{vy}/a mature sperm DNA. The figure is replicated from Blewitt *et al.* 2006, where it appears with the following caption: "Each line represents an individual clone, theoretically from one cell, and each circle an individual CpG. Open circles indicate an unmethylated CpG, and closed circles a methylated CpG. Each block of lines represents the clones derived from the sperm of one adult male mouse [….]. The percentage of methylation in each dataset is shown (calculated from the number of methylated CpGs divided by the total CpGs sequenced, multiplied by 100). The position of each circle is representative of the relative location along the length of the PCR product. Any clones with greater than 5% non-CpG methylation were excluded from the dataset, and these clones made up less than 5% of all clones sequenced." These data was first published in Rakyan *et al.* 2003. **B.** Methyl-sensitive restriction digest and Southern transfer of yellow and pseudoagouti A^{vy}/a mature sperm DNA. This figure is also replicated from Blewitt *et al.* 2006, where it appears with the following caption: "DNA was prepared and samples digested with BamHI followed by MspI or its isoschizomer HpaII, transferred and hybridised with the agouti probe. The A^{vy} allele produces a 9-kb BamHI band, while the *a* allele produces a 3.3-kb band. Membranes were stripped and rehybridised with a murine a-globin probe to check for equal digestion within the tissue samples […]. These results represent experiments performed on sperm and tail DNA from seven yellow and five pseudoagouti males […]. Mature sperm were isolated from both epididymes of the male (each sample contained in the order of 106 to 107 spermatocytes). Sperm samples were checked by light microscopy and found to be greater than 95% spermatocytes. The methylation state of the tissues is indicated by the ratio of the 9-kb BamHI band to the 7-kb band remaining after HpaII digestion. The 9-kb band is marked by an asterisk."

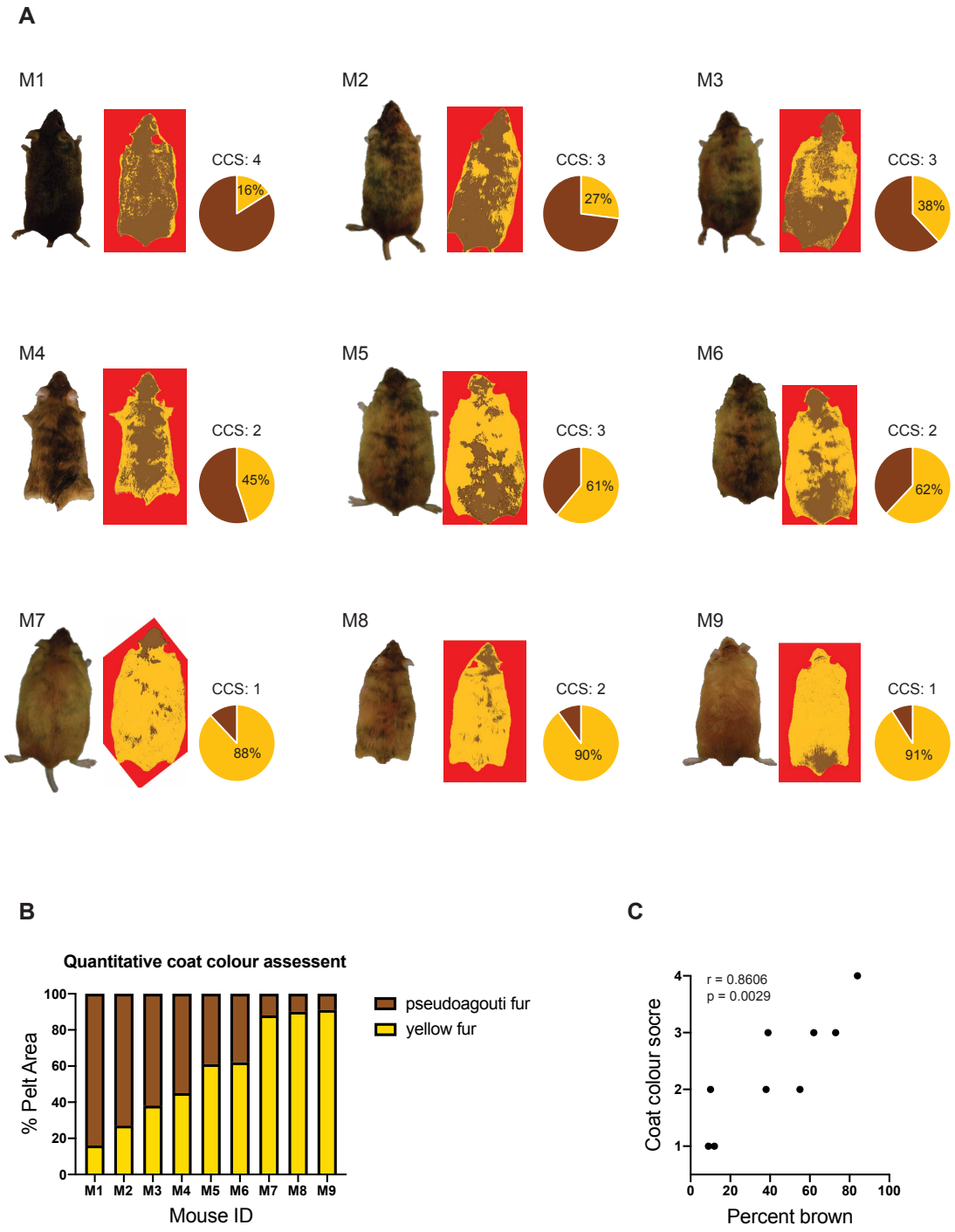


Figure A.3: Comparison of A^{vy} coat colour quantification methods. A. Nine A^{vy}/a males that span the A^{vy} coat colour spectrum were visually classified by David Schroeder using a subjective coat colour score (CCS) that ranged from 1 (yellow or nearly all yellow) to 4 (pseudoagouti or nearly all pseudoagouti). Coat colour was also measured computationally by Jennifer Stauffer using photographic input. Photographs were processed to produce areas of uniform yellow and brown colour (images with red backgrounds). The output data is reported as percent of pelt area that is yellow, shown as pie charts. Mice are labelled M1 through M9 and displayed in order of ascending percent yellow. **B.** Summary of quantitative coat colour assessment. **C.** The two coat colour quantification methods are highly correlated. The Pearson correlation coefficient (r) and p-value (p) are shown on the graph.

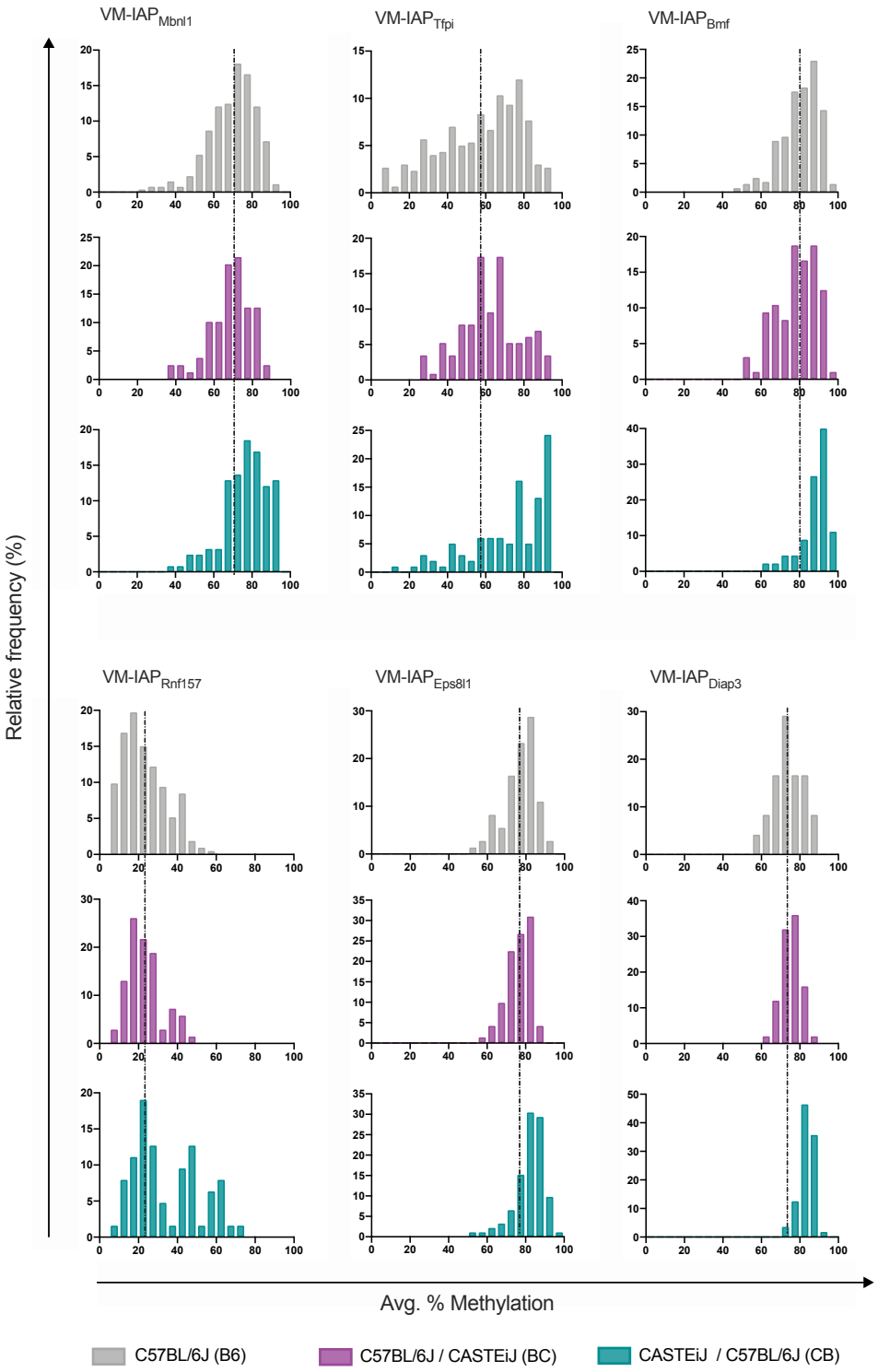


Figure A.4: VM-IAPs showing maternal genetic background effects (GBEs). Frequency distributions of VM-IAP methylation levels are shown for B6, BC, and CB populations. Methylation levels were averaged across the distal CpGs of the VM-IAP 5′ LTRs for each individual and classified into 20 bins (bin width = 5% methylation). Relative frequency was tabulated as a percentage of the population. Dashed vertical lines mark the mean VM-IAP methylation level in B6 mice and are included to facilitate the comparison of BC and CB distributions to B6 distributions.

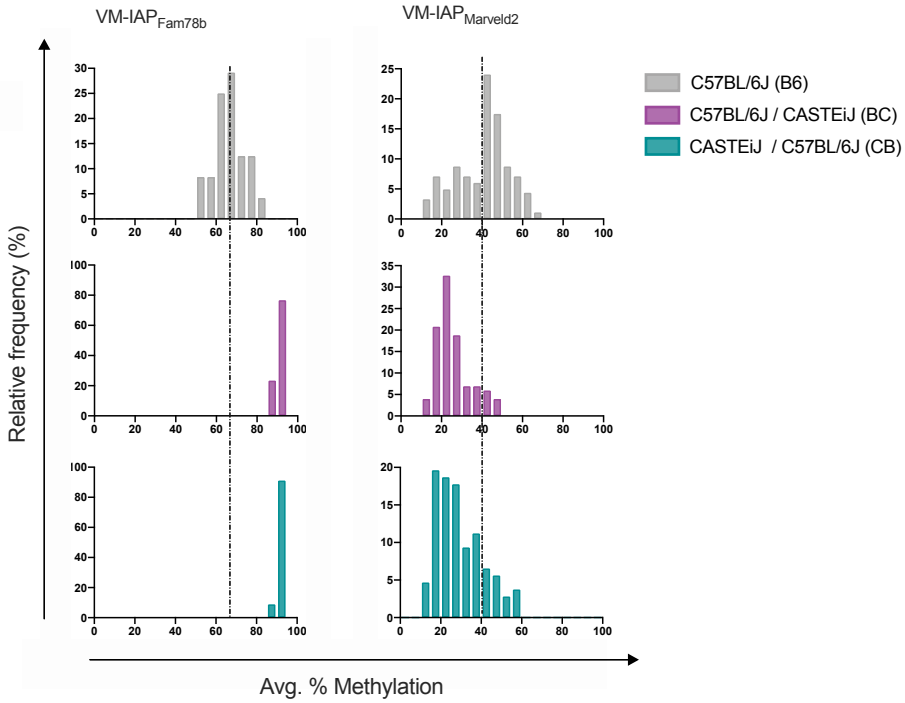


Figure A.5: VM-IAP$_{\text{Fam78b}}$ and VM-IAP$_{\text{Marveld2}}$ exhibit zygotic GBEs. Frequency distributions of VM-IAP methylation levels are shown for B6, BC, and CB populations. Methylation levels were averaged across the distal CpGs of the VM-IAP 5′ LTRs for each individual and classified into 20 bins (bin width = 5% methylation). Relative frequency was tabulated as a percentage of the population. Dashed vertical lines mark the mean VM-IAP methylation level in B6 mice and are included to facilitate the comparison of BC and CB distributions to B6 distributions.

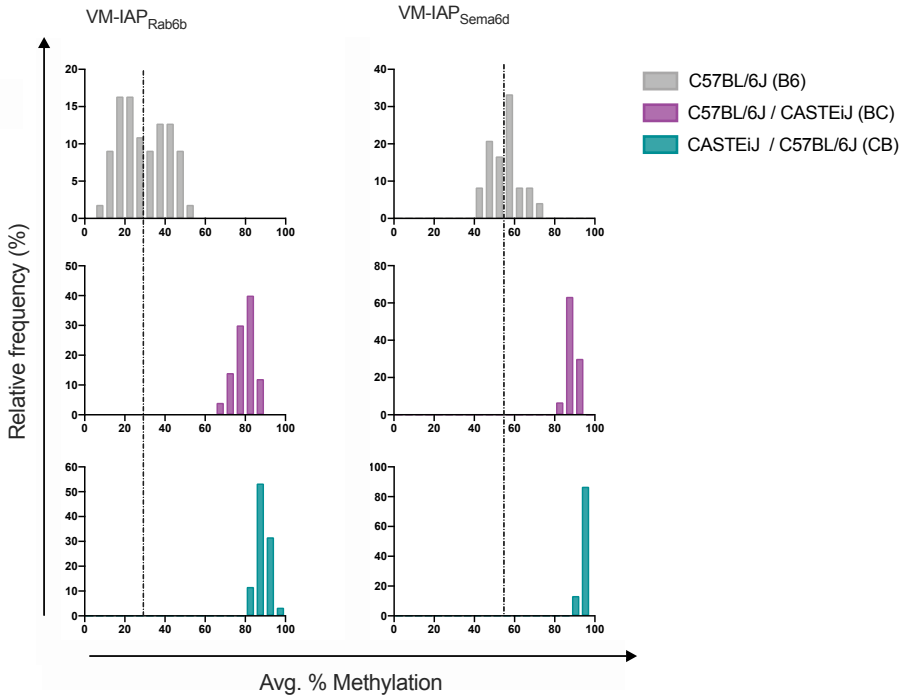


Figure A.6: VM-IAP$_{Rab6b}$ and VM-IAP$_{Sema6b}$ exhibit maternal and zygotic GBEs. Frequency distributions of VM-IAP methylation levels are shown for B6, BC, and CB populations. Methylation levels were averaged across the distal CpGs of the VM-IAP 5′ LTRs for each individual and classified into 20 bins (bin width = 5% methylation). Relative frequency was tabulated as a percentage of the population. Dashed vertical lines mark the mean VM-IAP methylation level in B6 mice and are included to facilitate the comparison of BC and CB distributions to B6 distributions.

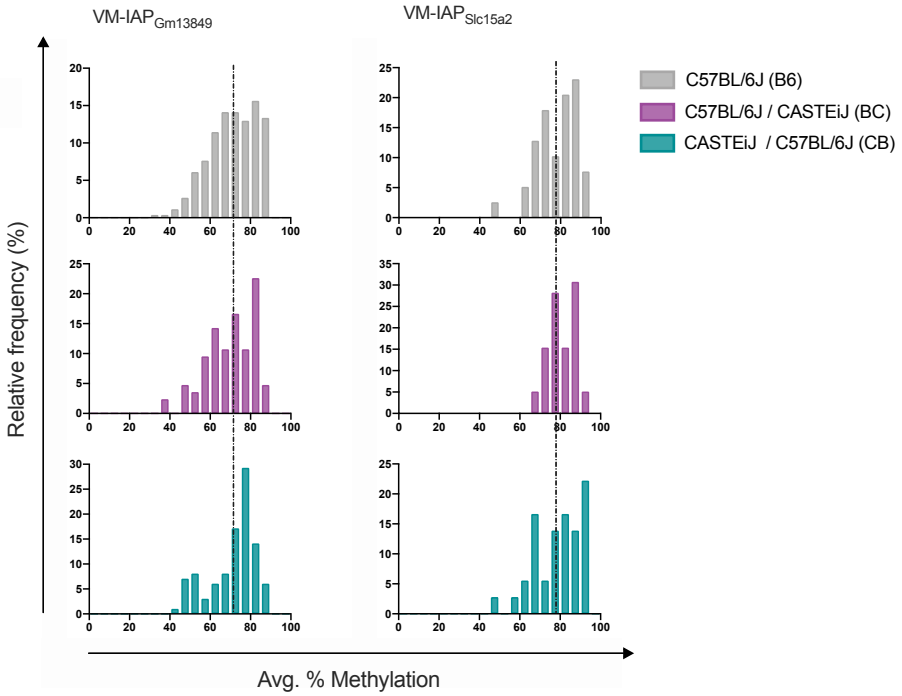
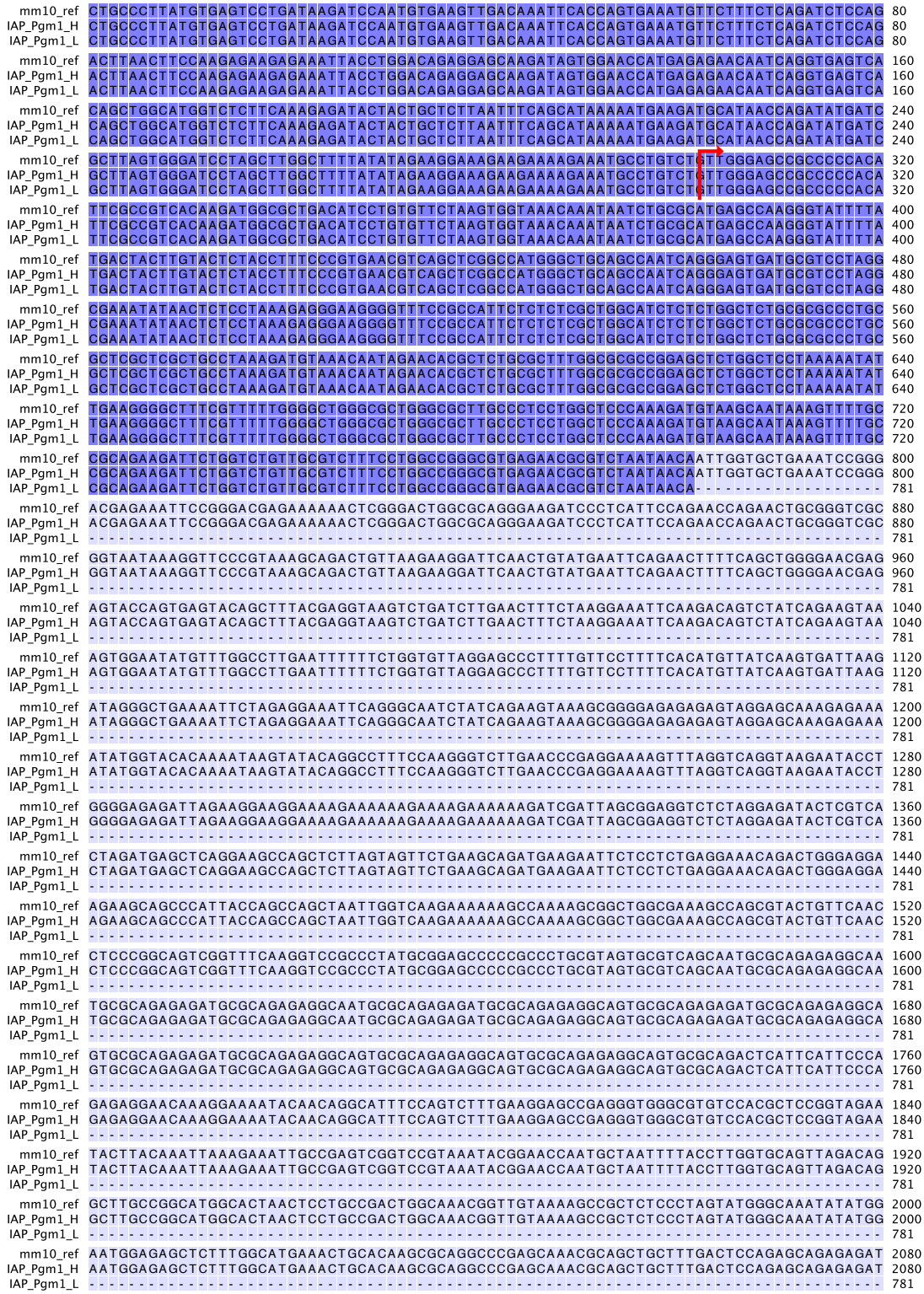
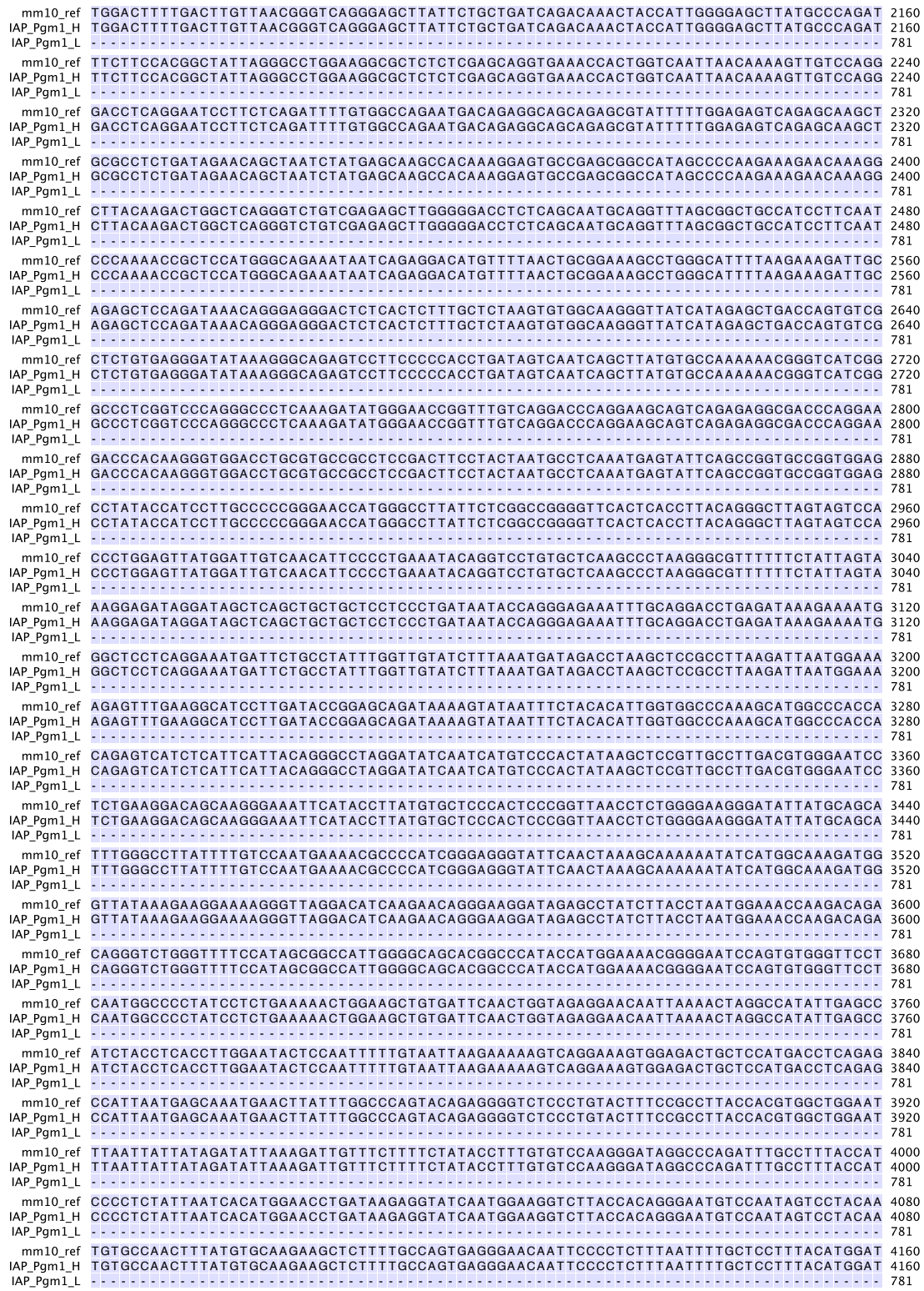
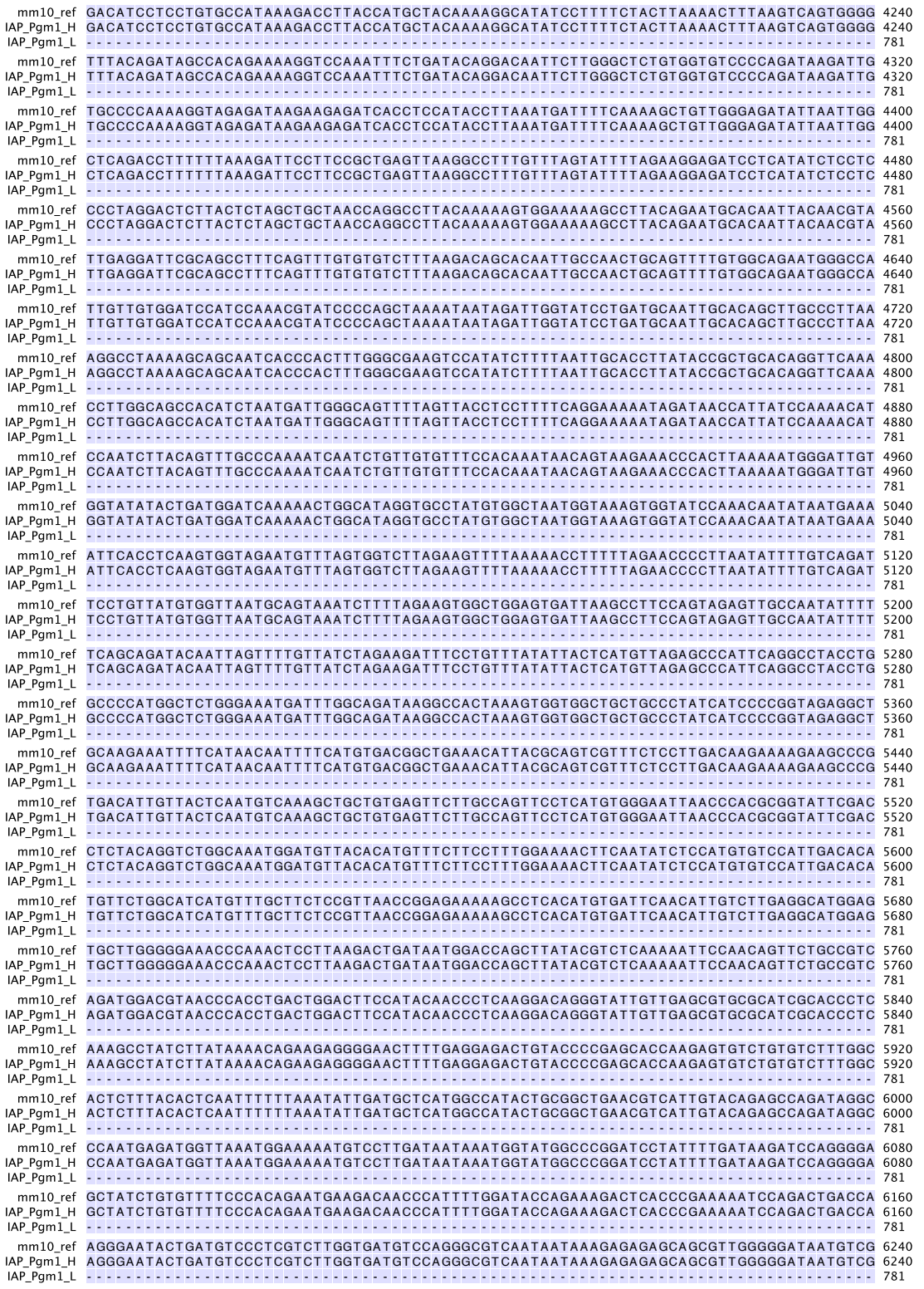


Figure A.7: VM-IAP$_{Gm13849}$ and VM-IAP$_{Slc15a2}$ show neither maternal nor zygotic GBEs. Frequency distributions of VM-IAP methylation levels are shown for B6 , BC, and CB populations. Methylation levels were averaged across the distal CpGs of the VM-IAP 5′ LTRs for each individual and classified into 20 bins (bin width = 5 % methylation). Relative frequency was tabulated as a percentage of the population. Dashed vertical lines mark the mean VM-IAP methylation level in B6 mice and are included to facilitate the comparison of BC and CB distributions to B6 distributions.


mm10_ref
IAP_Pgm1_H
IAP_Pgm1_L


mm10_ref TGGACTTTTGACTTGTTAACGGGTCAGGGAGCTTATTCTGCTGATCAGACAAACTACCATTGGGGAGCTTATGCCCAGAT 2160
IAP_Pgm1_H TGGACTTTTGACTTGTTAACGGGTCAGGGAGCTTATTCTGCTGATCAGACAAACTACCATTGGGGAGCTTATGCCCAGAT 2160
IAP_Pgm1_L ---------- 781
mm10_ref TTCTTCCACGGCTATTAGGGCCTGGAAGGCGCTCTCTCGAGCAGGTGAAACCACTGGTCAATTAACAAAAGTTGTCCAGG 2240
IAP_Pgm1_H TTCTTCCACGGCTATTAGGGCCTGGAAGGCGCTCTCTCGAGCAGGTGAAACCACTGGTCAATTAACAAAAGTTGTCCAGG 2240
IAP_Pgm1_L ---------- 781
mm10_ref GACCTCAGGAATCCTTCTCAGATTTTGTGGCCAGAATGACAGAGGCAGCAGAGCGTATTTTTGGAGAGTCAGAGCAAGCT 2320
IAP_Pgm1_H GACCTCAGGAATCCTTCTCAGATTTTGTGGCCAGAATGACAGAGGCAGCAGAGCGTATTTTTGGAGAGTCAGAGCAAGCT 2320
IAP_Pgm1_L ---------- 781
mm10_ref GCGCCTCTGATAGAACAGCTAATCTATGAGCAAGCCACAAAGGAGTGCCGAGCGGCCATAGCCCCAAGAAAGAACAAAGG 2400
IAP_Pgm1_H GCGCCTCTGATAGAACAGCTAATCTATGAGCAAGCCACAAAGGAGTGCCGAGCGGCCATAGCCCCAAGAAAGAACAAAGG 2400
IAP_Pgm1_L ---------- 781
mm10_ref CTTACAAGACTGGCTCAGGGTCTGTCGAGAGCTTGGGGGACCTCTCAGCAATGCAGGTTTAGCGGCTGCCATCCTTCAAT 2480
IAP_Pgm1_H CTTACAAGACTGGCTCAGGGTCTGTCGAGAGCTTGGGGGACCTCTCAGCAATGCAGGTTTAGCGGCTGCCATCCTTCAAT 2480
IAP_Pgm1_L ---------- 781
mm10_ref CCCAAAACCGCTCCATGGGCAGAAATAATCAGAGGACATGTTTTAACTGCGGAAAGCCTGGGCATTTTAAGAAAGATTGC 2560
IAP_Pgm1_H CCCAAAACCGCTCCATGGGCAGAAATAATCAGAGGACATGTTTTAACTGCGGAAAGCCTGGGCATTTTAAGAAAGATTGC 2560
IAP_Pgm1_L ---------- 781
mm10_ref AGAGCTCCAGATAAACAGGGAGGGACTCTCACTCTTTGCTCTAAGTGTGGCAAGGGTTATCATAGAGCTGACCAGTGTCG 2640
IAP_Pgm1_H AGAGCTCCAGATAAACAGGGAGGGACTCTCACTCTTTGCTCTAAGTGTGGCAAGGGTTATCATAGAGCTGACCAGTGTCG 2640
IAP_Pgm1_L ---------- 781
mm10_ref CTCTGTGAGGGATATAAAGGGCAGAGTCCTTCCCCCACCTGATAGTCAATCAGCTTATGTGCCAAAAAACGGGTCATCGG 2720
IAP_Pgm1_H CTCTGTGAGGGATATAAAGGGCAGAGTCCTTCCCCCACCTGATAGTCAATCAGCTTATGTGCCAAAAAACGGGTCATCGG 2720
IAP_Pgm1_L ---------- 781
mm10_ref GCCCTCGGTCCCAGGGCCCTCAAAGATATGGGAACCGGTTTGTCAGGACCCAGGAAGCAGTCAGAGAGGCGACCCAGGAA 2800
IAP_Pgm1_H GCCCTCGGTCCCAGGGCCCTCAAAGATATGGGAACCGGTTTGTCAGGACCCAGGAAGCAGTCAGAGAGGCGACCCAGGAA 2800
IAP_Pgm1_L ---------- 781
mm10_ref GACCCACAAGGGTGGACCTGCGTGCCGCCTCCGACTTCCTACTAATGCCTCAAATGAGTATTCAGCCGGTGCCGGTGGAG 2880
IAP_Pgm1_H GACCCACAAGGGTGGACCTGCGTGCCGCCTCCGACTTCCTACTAATGCCTCAAATGAGTATTCAGCCGGTGCCGGTGGAG 2880
IAP_Pgm1_L ---------- 781
mm10_ref CCTATACCATCCTTGCCCCCGGGAACCATGGGCCTTATTCTCGGCCGGGGTTCACTCACCTTACAGGGCTTAGTAGTCCA 2960
IAP_Pgm1_H CCTATACCATCCTTGCCCCCGGGAACCATGGGCCTTATTCTCGGCCGGGGTTCACTCACCTTACAGGGCTTAGTAGTCCA 2960
IAP_Pgm1_L ---------- 781
mm10_ref CCCTGGAGTTATGGATTGTCAACATTCCCCTGAAATACAGGTCCTGTGCTCAAGCCCTAAGGGCGTTTTTTCTATTAGTA 3040
IAP_Pgm1_H CCCTGGAGTTATGGATTGTCAACATTCCCCTGAAATACAGGTCCTGTGCTCAAGCCCTAAGGGCGTTTTTTCTATTAGTA 3040
IAP_Pgm1_L ---------- 781
mm10_ref AAGGAGATAGGATAGCTCAGCTGCTGCTCCTCCCTGATAATACCAGGGAGAAATTTGCAGGACCTGAGATAAAGAAAATG 3120
IAP_Pgm1_H AAGGAGATAGGATAGCTCAGCTGCTGCTCCTCCCTGATAATACCAGGGAGAAATTTGCAGGACCTGAGATAAAGAAAATG 3120
IAP_Pgm1_L ---------- 781
mm10_ref GGCTCCTCAGGAAATGATTCTGCCTATTTGGTTGTATCTTTAAATGATAGACCTAAGCTCCGCCTTAAGATTAATGGAAA 3200
IAP_Pgm1_H GGCTCCTCAGGAAATGATTCTGCCTATTTGGTTGTATCTTTAAATGATAGACCTAAGCTCCGCCTTAAGATTAATGGAAA 3200
IAP_Pgm1_L ---------- 781
mm10_ref AGAGTTTGAAGGCATCCTTGATACCGGAGCAGATAAAAGTATAATTTCTACACATTGGTGGCCCAAAGCATGGCCCACCA 3280
IAP_Pgm1_H AGAGTTTGAAGGCATCCTTGATACCGGAGCAGATAAAAGTATAATTTCTACACATTGGTGGCCCAAAGCATGGCCCACCA 3280
IAP_Pgm1_L ---------- 781
mm10_ref CAGAGTCATCTCATTCATTACAGGGCCTAGGATATCAATCATGTCCCACTATAAGCTCCGTTGCCTTGACGTGGGAATCC 3360
IAP_Pgm1_H CAGAGTCATCTCATTCATTACAGGGCCTAGGATATCAATCATGTCCCACTATAAGCTCCGTTGCCTTGACGTGGGAATCC 3360
IAP_Pgm1_L ---------- 781
mm10_ref TCTGAAGGACAGCAAGGGAAATTCATACCTTATGTGCTCCCACTCCCGGTTAACCTCTGGGGAAGGGATATTATGCAGCA 3440
IAP_Pgm1_H TCTGAAGGACAGCAAGGGAAATTCATACCTTATGTGCTCCCACTCCCGGTTAACCTCTGGGGAAGGGATATTATGCAGCA 3440
IAP_Pgm1_L ---------- 781
mm10_ref TTTGGGCCTTATTTTGTCCAATGAAAACGCCCCATCGGGAGGGTATTCAACTAAAGCAAAAAATATCATGGCAAAGATGG 3520
IAP_Pgm1_H TTTGGGCCTTATTTTGTCCAATGAAAACGCCCCATCGGGAGGGTATTCAACTAAAGCAAAAAATATCATGGCAAAGATGG 3520
IAP_Pgm1_L ---------- 781
mm10_ref GTTATAAAGAAGGAAAAGGGTTAGGACATCAAGAACAGGGAAGGATAGAGCCTATCTTACCTAATGGAAACCAAGACAGA 3600
IAP_Pgm1_H GTTATAAAGAAGGAAAAGGGTTAGGACATCAAGAACAGGGAAGGATAGAGCCTATCTTACCTAATGGAAACCAAGACAGA 3600
IAP_Pgm1_L ---------- 781
mm10_ref CAGGGTCTGGGTTTTCCATAGCGGCCATTGGGGCAGCACGGCCCATACCATGGAAAACGGGGAATCCAGTGTGGGTTCCT 3680
IAP_Pgm1_H CAGGGTCTGGGTTTTCCATAGCGGCCATTGGGGCAGCACGGCCCATACCATGGAAAACGGGGAATCCAGTGTGGGTTCCT 3680
IAP_Pgm1_L ---------- 781
mm10_ref CAATGGCCCCTATCCTCTGAAAAACTGGAAGCTGTGATTCAACTGGTAGAGGAACAATTAAAACTAGGCCATATTGAGCC 3760
IAP_Pgm1_H CAATGGCCCCTATCCTCTGAAAAACTGGAAGCTGTGATTCAACTGGTAGAGGAACAATTAAAACTAGGCCATATTGAGCC 3760
IAP_Pgm1_L ---------- 781
mm10_ref ATCTACCTCACCTTGGAATACTCCAATTTTTGTAATTAAGAAAAAGTCAGGAAAGTGGAGACTGCTCCATGACCTCAGAG 3840
IAP_Pgm1_H ATCTACCTCACCTTGGAATACTCCAATTTTTGTAATTAAGAAAAAGTCAGGAAAGTGGAGACTGCTCCATGACCTCAGAG 3840
IAP_Pgm1_L ---------- 781
mm10_ref CCATTAATGAGCAAATGAACTTATTTGGCCCAGTACAGAGGGGTCTCCCTGTACTTTCCGCCTTACCACGTGGCTGGAAT 3920
IAP_Pgm1_H CCATTAATGAGCAAATGAACTTATTTGGCCCAGTACAGAGGGGTCTCCCTGTACTTTCCGCCTTACCACGTGGCTGGAAT 3920
IAP_Pgm1_L ---------- 781
mm10_ref TTAATTATTATAGATATTAAAGATTGTTTCTTTTCTATACCTTTGTGTCCAAGGGATAGGCCCAGATTTGCCTTTACCAT 4000
IAP_Pgm1_H TTAATTATTATAGATATTAAAGATTGTTTCTTTTCTATACCTTTGTGTCCAAGGGATAGGCCCAGATTTGCCTTTACCAT 4000
IAP_Pgm1_L ---------- 781
mm10_ref CCCCTCTATTAATCACATGGAACCTGATAAGAGGTATCAATGGAAGGTCTTACCACAGGGAATGTCCAATAGTCCTACAA 4080
IAP_Pgm1_H CCCCTCTATTAATCACATGGAACCTGATAAGAGGTATCAATGGAAGGTCTTACCACAGGGAATGTCCAATAGTCCTACAA 4080
IAP_Pgm1_L ---------- 781
mm10_ref TGTGCCAACTTTATGTGCAAGAAGCTCTTTTGCCAGTGAGGGAACAATTCCCCTCTTTAATTTTGCTCCTTTACATGGAT 4160
IAP_Pgm1_H TGTGCCAACTTTATGTGCAAGAAGCTCTTTTGCCAGTGAGGGAACAATTCCCCTCTTTAATTTTGCTCCTTTACATGGAT 4160
IAP_Pgm1_L ---------- 781


mm10_ref GACATCCTCCTGTGCCATAAAGACCTTACCATGCTACAAAAGGCATATCCTTTTCTACTTAAAACTTTAAGTCAGTGGGG 4240
IAP_Pgm1_H GACATCCTCCTGTGCCATAAAGACCTTACCATGCTACAAAAGGCATATCCTTTTCTACTTAAAACTTTAAGTCAGTGGGG 4240
IAP_Pgm1_L ---------- 781
mm10_ref TTTACAGATAGCCACAGAAAAGGTCCAAATTTCTGATACAGGACAATTCTTGGGCTCTGTGGTGTCCCCAGATAAGATTG 4320
IAP_Pgm1_H TTTACAGATAGCCACAGAAAAGGTCCAAATTTCTGATACAGGACAATTCTTGGGCTCTGTGGTGTCCCCAGATAAGATTG 4320
IAP_Pgm1_L ---------- 781
mm10_ref TGCCCCAAAAGGTAGAGATAAGAAGAGATCACCTCCATACCTTAAATGATTTTCAAAAGCTGTTGGGAGATATTAATTGG 4400
IAP_Pgm1_H TGCCCCAAAAGGTAGAGATAAGAAGAGATCACCTCCATACCTTAAATGATTTTCAAAAGCTGTTGGGAGATATTAATTGG 4400
IAP_Pgm1_L ---------- 781
mm10_ref CTCAGACCTTTTTTAAAGATTCCTTCCGCTGAGTTAAGGCCTTTGTTTAGTATTTTAGAAGGAGATCCTCATATCTCCTC 4480
IAP_Pgm1_H CTCAGACCTTTTTTAAAGATTCCTTCCGCTGAGTTAAGGCCTTTGTTTAGTATTTTAGAAGGAGATCCTCATATCTCCTC 4480
IAP_Pgm1_L ---------- 781
mm10_ref CCCTAGGACTCTTACTCTAGCTGCTAACCAGGCCTTACAAAAAGTGGAAAAAGCCTTACAGAATGCACAATTACAACGTA 4560
IAP_Pgm1_H CCCTAGGACTCTTACTCTAGCTGCTAACCAGGCCTTACAAAAAGTGGAAAAAGCCTTACAGAATGCACAATTACAACGTA 4560
IAP_Pgm1_L ---------- 781
mm10_ref TTGAGGATTCGCAGCCTTTCAGTTTGTGTGTCTTTAAGACAGCACAATTGCCAACTGCAGTTTTGTGGCAGAATGGGCCA 4640
IAP_Pgm1_H TTGAGGATTCGCAGCCTTTCAGTTTGTGTGTCTTTAAGACAGCACAATTGCCAACTGCAGTTTTGTGGCAGAATGGGCCA 4640
IAP_Pgm1_L ---------- 781
mm10_ref TTGTTGTGGATCCATCCAAACGTATCCCCAGCTAAAATAATAGATTGGTATCCTGATGCAATTGCACAGCTTGCCCTTAA 4720
IAP_Pgm1_H TTGTTGTGGATCCATCCAAACGTATCCCCAGCTAAAATAATAGATTGGTATCCTGATGCAATTGCACAGCTTGCCCTTAA 4720
IAP_Pgm1_L ---------- 781
mm10_ref AGGCCTAAAAGCAGCAATCACCCACTTTGGGCGAAGTCCATATCTTTTAATTGCACCTTATACCGCTGCACAGGTTCAAA 4800
IAP_Pgm1_H AGGCCTAAAAGCAGCAATCACCCACTTTGGGCGAAGTCCATATCTTTTAATTGCACCTTATACCGCTGCACAGGTTCAAA 4800
IAP_Pgm1_L ---------- 781
mm10_ref CCTTGGCAGCCACATCTAATGATTGGGCAGTTTTAGTTACCTCCTTTTCAGGAAAAATAGATAACCATTATCCAAAACAT 4880
IAP_Pgm1_H CCTTGGCAGCCACATCTAATGATTGGGCAGTTTTAGTTACCTCCTTTTCAGGAAAAATAGATAACCATTATCCAAAACAT 4880
IAP_Pgm1_L ---------- 781
mm10_ref CCAATCTTACAGTTTGCCCAAAATCAATCTGTTGTGTTTCCACAAATAACAGTAAGAAACCCACTTAAAAATGGGATTGT 4960
IAP_Pgm1_H CCAATCTTACAGTTTGCCCAAAATCAATCTGTTGTGTTTCCACAAATAACAGTAAGAAACCCACTTAAAAATGGGATTGT 4960
IAP_Pgm1_L ---------- 781
mm10_ref GGTATATACTGATGGATCAAAAACTGGCATAGGTGCCTATGTGGCTAATGGTAAAGTGGTATCCAAACAATATAATGAAA 5040
IAP_Pgm1_H GGTATATACTGATGGATCAAAAACTGGCATAGGTGCCTATGTGGCTAATGGTAAAGTGGTATCCAAACAATATAATGAAA 5040
IAP_Pgm1_L ---------- 781
mm10_ref ATTCACCTCAAGTGGTAGAATGTTTAGTGGTCTTAGAAGTTTTAAAAACCTTTTTAGAACCCCTTAATATTTTGTCAGAT 5120
IAP_Pgm1_H ATTCACCTCAAGTGGTAGAATGTTTAGTGGTCTTAGAAGTTTTAAAAACCTTTTTAGAACCCCTTAATATTTTGTCAGAT 5120
IAP_Pgm1_L ---------- 781
mm10_ref TCCTGTTATGTGGTTAATGCAGTAAATCTTTTAGAAGTGGCTGGAGTGATTAAGCCTTCCAGTAGAGTTGCCAATATTTT 5200
IAP_Pgm1_H TCCTGTTATGTGGTTAATGCAGTAAATCTTTTAGAAGTGGCTGGAGTGATTAAGCCTTCCAGTAGAGTTGCCAATATTTT 5200
IAP_Pgm1_L ---------- 781
mm10_ref TCAGCAGATACAATTAGTTTTGTTATCTAGAAGATTTCCTGTTTATATTACTCATGTTAGAGCCCATTCAGGCCTACCTG 5280
IAP_Pgm1_H TCAGCAGATACAATTAGTTTTGTTATCTAGAAGATTTCCTGTTTATATTACTCATGTTAGAGCCCATTCAGGCCTACCTG 5280
IAP_Pgm1_L ---------- 781
mm10_ref GCCCCATGGCTCTGGGAAATGATTTGGCAGATAAGGCCACTAAAGTGGTGGCTGCTGCCCTATCATCCCCGGTAGAGGCT 5360
IAP_Pgm1_H GCCCCATGGCTCTGGGAAATGATTTGGCAGATAAGGCCACTAAAGTGGTGGCTGCTGCCCTATCATCCCCGGTAGAGGCT 5360
IAP_Pgm1_L ---------- 781
mm10_ref GCAAGAAATTTTCATAACAATTTTCATGTGACGGCTGAAACATTACGCAGTCGTTTCTCCTTGACAAGAAAAGAAGCCCG 5440
IAP_Pgm1_H GCAAGAAATTTTCATAACAATTTTCATGTGACGGCTGAAACATTACGCAGTCGTTTCTCCTTGACAAGAAAAGAAGCCCG 5440
IAP_Pgm1_L ---------- 781
mm10_ref TGACATTGTTACTCAATGTCAAAGCTGCTGTGAGTTCTTGCCAGTTCCTCATGTGGGAATTAACCCACGCGGTATTCGAC 5520
IAP_Pgm1_H TGACATTGTTACTCAATGTCAAAGCTGCTGTGAGTTCTTGCCAGTTCCTCATGTGGGAATTAACCCACGCGGTATTCGAC 5520
IAP_Pgm1_L ---------- 781
mm10_ref CTCTACAGGTCTGGCAAATGGATGTTACACATGTTTCTTCCTTTGGAAAACTTCAATATCTCCATGTGTCCATTGACACA 5600
IAP_Pgm1_H CTCTACAGGTCTGGCAAATGGATGTTACACATGTTTCTTCCTTTGGAAAACTTCAATATCTCCATGTGTCCATTGACACA 5600
IAP_Pgm1_L ---------- 781
mm10_ref TGTTCTGGCATCATGTTTGCTTCTCCGTTAACCGGAGAAAAAGCCTCACATGTGATTCAACATTGTCTTGAGGCATGGAG 5680
IAP_Pgm1_H TGTTCTGGCATCATGTTTGCTTCTCCGTTAACCGGAGAAAAAGCCTCACATGTGATTCAACATTGTCTTGAGGCATGGAG 5680
IAP_Pgm1_L ---------- 781
mm10_ref TGCTTGGGGGAAACCCAAACTCCTTAAGACTGATAATGGACCAGCTTATACGTCTCAAAAATTCCAACAGTTCTGCCGTC 5760
IAP_Pgm1_H TGCTTGGGGGAAACCCAAACTCCTTAAGACTGATAATGGACCAGCTTATACGTCTCAAAAATTCCAACAGTTCTGCCGTC 5760
IAP_Pgm1_L ---------- 781
mm10_ref AGATGGACGTAACCCACCTGACTGGACTTCCATACAACCCTCAAGGACAGGGTATTGTTGAGCGTGCGCATCGCACCCTC 5840
IAP_Pgm1_H AGATGGACGTAACCCACCTGACTGGACTTCCATACAACCCTCAAGGACAGGGTATTGTTGAGCGTGCGCATCGCACCCTC 5840
IAP_Pgm1_L ---------- 781
mm10_ref AAAGCCTATCTTATAAAACAGAAGAGGGGAACTTTTGAGGAGACTGTACCCCGAGCACCAAGAGTGTCTGTGTCTTTGGC 5920
IAP_Pgm1_H AAAGCCTATCTTATAAAACAGAAGAGGGGAACTTTTGAGGAGACTGTACCCCGAGCACCAAGAGTGTCTGTGTCTTTGGC 5920
IAP_Pgm1_L ---------- 781
mm10_ref ACTCTTTACACTCAATTTTTTAAATATTGATGCTCATGGCCATACTGCGGCTGAACGTCATTGTACAGAGCCAGATAGGC 6000
IAP_Pgm1_H ACTCTTTACACTCAATTTTTTAAATATTGATGCTCATGGCCATACTGCGGCTGAACGTCATTGTACAGAGCCAGATAGGC 6000
IAP_Pgm1_L ---------- 781
mm10_ref CCAATGAGATGGTTAAATGGAAAAATGTCCTTGATAATAAATGGTATGGCCCGGATCCTATTTTGATAAGATCCAGGGGA 6080
IAP_Pgm1_H CCAATGAGATGGTTAAATGGAAAAATGTCCTTGATAATAAATGGTATGGCCCGGATCCTATTTTGATAAGATCCAGGGGA 6080
IAP_Pgm1_L ---------- 781
mm10_ref GCTATCTGTGTTTTCCCACAGAATGAAGACAACCCATTTTGGATACCAGAAAGACTCACCCGAAAAATCCAGACTGACCA 6160
IAP_Pgm1_H GCTATCTGTGTTTTCCCACAGAATGAAGACAACCCATTTTGGATACCAGAAAGACTCACCCGAAAAATCCAGACTGACCA 6160
IAP_Pgm1_L ---------- 781
mm10_ref AGGGAATACTGATGTCCCTCGTCTTGGTGATGTCCAGGGCGTCAATAATAAAGAGAGAGCAGCGTTGGGGGATAATGTCG 6240
IAP_Pgm1_H AGGGAATACTGATGTCCCTCGTCTTGGTGATGTCCAGGGCGTCAATAATAAAGAGAGAGCAGCGTTGGGGGATAATGTCG 6240
IAP_Pgm1_L ---------- 781

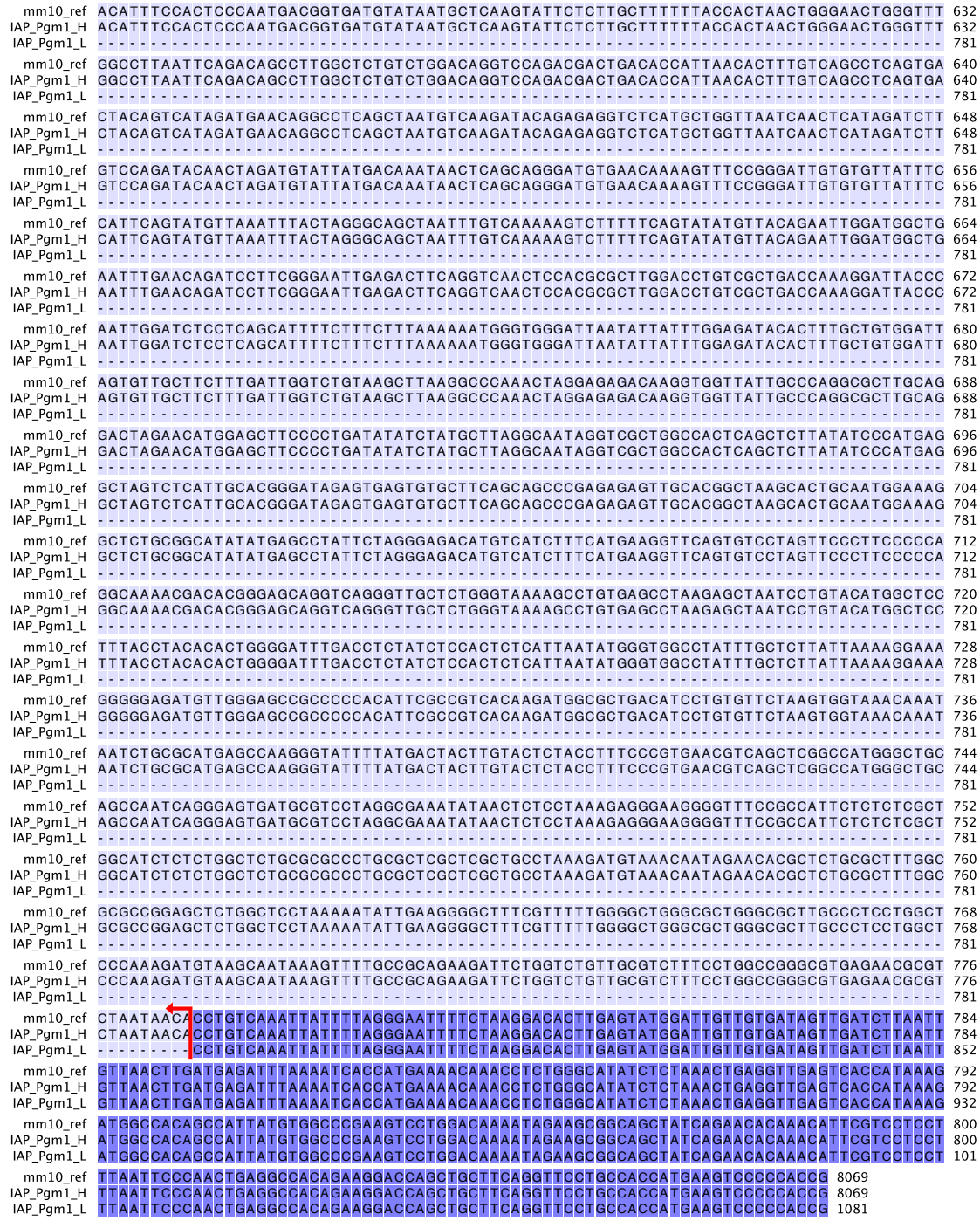


Figure A.8: Base-resolution alignment of the IAP$_{Pgm1^H}$ and IAP$_{Pgm1^L}$ alleles. The full IAP$_{Pgm1^H}$ (IAP_Pgm1_H) and IAP$_{Pgm1^L}$ (IAP_Pgm1_L) sequences were assembled from Sanger sequencing reads and aligned to the GRC38/mm10 reference sequence for IAP$_{Pgm1}$ (mm10_ref). Perfect consensus across all three sequences is shown in dark blue. Perfect consensus between the reference and the IAP$_{Pgm1^H}$ sequence is shown in light blue. Dashes indicate lack of sequence. The alignment of the IAP$_{Pgm1^L}$ solo LTR to the 5′ LTR of the IAP$_{Pgm1^H}$ allele is arbitrary and could equivalently be aligned to the 3′ LTR. The beginning and end of the IAP element are shown with red arrows. The alignment was generated with CLC Sequence Viewer 6.

Appendix B

Related publications

Kazachenka, A.*, Bertozzi, T.M.*, Sjoberg-Herrera, M.K., Walker, N., Gardner, J., Gunning, R., Pahita, E., Adams, S., Adams, D. & Ferguson-Smith, A.C. (2018). Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. *Cell*, **175**(5), 1259-1271. **co-first authors*

This paper contains data presented in Chapters 2 and 3 of this dissertation.

Bertozzi, T.M. & Ferguson-Smith, A.C. (2019). Metastable epialleles and their contribution to epigenetic inheritance in mammals. *Seminars in Cell and Developmental Biology*. doi: 10.1016/J.SEMCDB.2019.08.002

This review largely consists of the second half of the introduction to this dissertation.

Bertozzi, T.M., Blake, G.E.T., Bansal A., Nguyen, D.K., Fernandez-Twinn, D.S., Ozanne, S.E., Bartolomei, M.S., Simmons, R.A., Watson, E.D., Ferguson-Smith, A.C. Variably methylated epialleles are responsive to perturbed folate metabolism but are insensitive to other environmental contexts. *Manuscript under revision.*

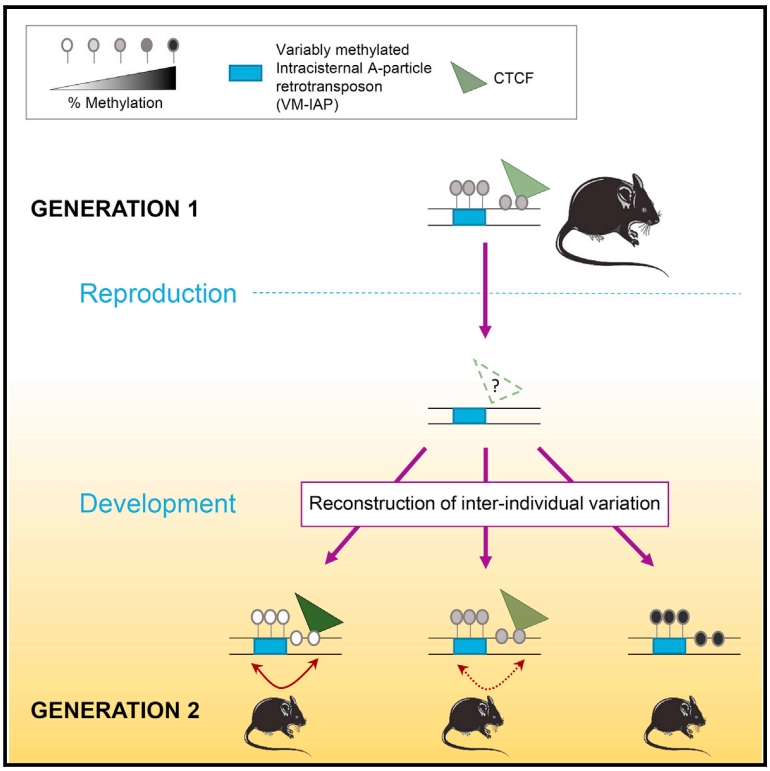
This manuscript contains most of the results presented in Chapter 4 of this dissertation.

Article

Cell

Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance

Graphical Abstract



Highlights

- Repertoire of variably methylated repeat elements defined in inbred mice
- VM-IAPs are flanked by CTCF binding sites, and very few act as promoters
- Methylation variability is re-established from one generation to the next
- Memory of parental methylation state is an exception rather than the rule

Authors

Anastasiya Kazachenka, Tessa M. Bertozzi, Marcela K. Sjoberg-Herrera, ..., Sarah Adams, David Adams, Anne C. Ferguson-Smith

Correspondence

afsmith@gen.cam.ac.uk

In Brief

A genome-wide interrogation of variably methylated endogenous retroviruses shows common reprogramming of methylation states after fertilization, challenging the paradigm of transgenerational non-genetic inheritance at such loci.

Kazachenka et al., 2018, Cell *175*, 1259–1271
November 15, 2018
https://doi.org/10.1016/j.cell.2018.09.043

CellPress

Article

Cell

Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance

Anastasiya Kazachenka,[1,3] Tessa M. Bertozzi,[1,3] Marcela K. Sjoberg-Herrera,[2,4] Nic Walker,[1,5] Joseph Gardner,[1] Richard Gunning,[2] Elena Pahita,[1] Sarah Adams,[1] David Adams,[2] and Anne C. Ferguson-Smith[1,6,*]

[1]Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK

[2]Experimental Cancer Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK

[3]These authors contributed equally

[4]Present address: Departamento de Biología Celular y Molecular, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile

[5]Present address: Cambridge Epigenetix, The Trinity Building, Chesterford Business Park, Little Chesterford, Saffron Walden CB10 1XL, UK

[6]Lead Contact

*Correspondence: afsmith@gen.cam.ac.uk

https://doi.org/10.1016/j.cell.2018.09.043

SUMMARY

Generally repressed by epigenetic mechanisms, retrotransposons represent around 40% of the murine genome. At the *Agouti viable yellow* (A^{vy}) locus, an endogenous retrovirus (ERV) of the intracisternal A particle (IAP) class retrotransposed upstream of the *agouti* coat-color locus, providing an alternative promoter that is variably DNA methylated in genetically identical individuals. This results in variable expressivity of coat color that is inherited transgenerationally. Here, a systematic genome-wide screen identifies multiple C57BL/6J murine IAPs with A^{vy} epigenetic properties. Each exhibits a stable methylation state within an individual but varies between individuals. Only in rare instances do they act as promoters controlling adjacent gene expression. Their methylation state is locus-specific within an individual, and their flanking regions are enriched for CTCF. Variably methylated IAPs are reprogrammed after fertilization and re-established as variable loci in the next generation, indicating reconstruction of metastable epigenetic states and challenging the generalizability of non-genetic inheritance at these regions.

INTRODUCTION

Most interindividual phenotypic variation is explained by genetic variation. However, studies in plant and animal models indicate that non-genetic mechanisms can contribute to phenotypic variability, and such phenotypes can be inherited over multiple generations (Cubas et al., 1999; Morgan and Whitelaw, 2008; Becker and Weigel, 2012). Epigenetic changes in the absence of genetic effects have been reported to have long-lasting phenotypic outcomes over multiple generations in non-mammalian organisms. In mammals, such non-genetic effects are difficult to explain mechanistically, and it has been challenging to define the regulatory processes underlying the observed phenomena (Miska and Ferguson-Smith, 2016).

Two of the best-characterized paradigms of non-genetic inheritance in mammals occur at the murine *Agouti viable yellow* (A^{vy}) and *Axin Fused* ($Axin^{Fu}$) loci (Dickies, 1962; Vasicek et al., 1997). In these naturally occurring mutant mice, genetically identical individuals exhibit quantifiable phenotypic variability in coat color or tail morphology due to the insertion of an endogenous retrovirus (ERV) of the intracisternal A particle (IAP) class into the *Agouti* or the *Fused* loci, respectively. The range of phenotypes correlates reproducibly with interindividual differences in the level of DNA methylation at a long terminal repeat (LTR) promoter of the IAP, driving abnormal expression of the genes (Michaud et al., 1994; Rakyan et al., 2003). The consistency in methylation level observed within an individual is in contrast to the variation of methylation levels and phenotypic outcomes observed between individuals, defining A^{vy} and $Axin^{Fu}$ as so-called "metastable epialleles" (Rakyan et al., 2002). Transgenerational inheritance of the methylation pattern at these metastable epialleles has been observed, whereby the distribution of phenotypes in the offspring was shown to be dependent on parental phenotype (Morgan et al., 1999; Rakyan et al., 2003). Furthermore, A^{vy} is susceptible to environmental influence impacting methylation and phenotype (Wolff et al., 1998; Dolinoy et al., 2006, 2007; Kaminen-Ahola et al., 2010). Using genetic screens, proteins with epigenetic function associated with the maintenance of A^{vy} have been identified (Daxinger et al., 2013). In another study, a C57BL/6J endogenous IAP insertion at *Cdk5rap1* regulates transcriptional dosage via promoter methylation; however, an association with phenotype has not been reported (Druker et al., 2004). Together, these studies suggest that ERVs of the IAP subclass have the potential to be variably methylated, here referred to as variably methylated IAPs (VM-IAPs).

The properties and underlying mechanisms governing the establishment, behavior, and inheritance of VM-IAPs remain elusive, as does the extent to which they represent a genome-wide phenomenon. 45% of the murine genome is made up of

repetitive sequences, with ERVs comprising about 12% of the genome. In the C57BL/6J genome, there are approximately 12,000 ERVs of the IAP subclass (Smit et al., 2015). The degree to which this substantial fraction of the repeat genome might modulate phenotype is unclear, and the total number of naturally existing murine VM-IAPs is unknown to date.

Previous studies have searched for metastable epialleles with limited success. Strategies have included surveying expression microarray data for within-strain interindividual expression patterns, screening for retrotransposons that neighbor promoters marked by the active histone modification H3K4me3, and conducting a phylogenetic analysis on IAP elements (Weinhouse et al., 2011; Ekram et al., 2012; Faulk et al., 2013). A recent more extensive screen used comparative whole-genome bisulfite sequencing (WGBS) data and described 55 ERV regions exhibiting some interindividual differential methylation, with validation in two tissues shown for four (Oey et al., 2015). This study confirmed that naturally occurring germline mutations and interindividual genetic differences do not underlie the epigenetic variation observed at the identified regions. While individually informative, there is little or no overlap between the results of these screens. The more challenging task of identifying human metastable epialleles has been tackled before, but the genetic heterogeneity associated with human cohorts remains a significant hindrance in such studies (Silver et al., 2015).

Here, we report a novel high-stringency genome-wide approach to comprehensively identify VM-IAPs. Using WGBS and RNA sequencing (RNA-seq) datasets generated from pure non-cycling populations of *ex vivo* purified naive B and T cells, we identified individual elements possessing features of metastable epialleles. After extensive validation, we have characterized their relationship to each other and to the vast majority of IAPs in the genome that are fully and stably modified. Furthermore, we determined their patterns of inheritance from one generation to the next. Our study identifies a repertoire of loci with the potential to act as markers of normal and compromised environmental contexts and as tools to uncover mechanisms of non-genetic inheritance and, more generally, to provide insights into the mechanisms of silencing at repeats and the impact of mammalian repetitive elements on genome function and phenotype.

RESULTS

Identification of the VM-IAP Methylation Pattern

A three-step approach was used to identify VM-IAPs with metastable epiallele properties genome-wide. Our starting point for defining a metastable epiallele was (a) interindividual methylation variation at the IAP-LTR promoter, (b) consistent intraindividual methylation, and (c) variation in expression at an adjacent gene, as described for previously identified metastable epialleles. To this end, the first step utilized a catalog of polymorphic IAPs "private" to C57BL/6J compared to CAST/EiJ mice and screened for C57BL/6J-specific IAPs potentially impacting the expression of neighboring genes. The second step used the set of identified VM-IAP candidates from the first step to develop an algorithm to identify VM-IAPs genome-wide, which was applied to all IAPs in the C57BL/6J genome regardless of impact on adjacent expression. The third step consisted of running the algorithm on all C57BL/6J ERVs to assess the extent to which other ERV subclasses can act as metastable epialleles.

According to published data, 1994 IAP insertions are present in the C57BL/6J mouse strain and absent from the CAST/Eij strain (Nellåker et al., 2012). For the first step of our screen, we hypothesized that such polymorphic IAP insertions could explain some of the differential gene expression observed between the two strains. Differentially expressed genes were identified using RNA-seq datasets generated from naive non-cycling B and T cell populations purified from C57BL/6J and CAST/Eij mice generated as part of the BLUEPRINT reference epigenome project (Adams et al., 2012; accession number: GSE94676). The 552 polymorphic IAPs lying within or near differentially expressed genes were selected as potential metastable epialleles.

Methylation profiles of the identified 552 C57BL/6J IAPs were extracted from datasets generated from the same B and T cell populations used for the RNA-seq datasets. Both cell types were used because the methylation level at a metastable epiallele is established early in development and is therefore consistent within the same individual (Dolinoy et al., 2006; Waterland et al., 2006). As expected, the vast majority of these IAPs were highly methylated across all datasets. However, 31 showed a distinct methylation pattern at the IAP LTR characterized by "ragged" methylation levels between replicates (Figure 1A).

To test the hypothesis that ragged methylation reflected interindividual methylation variation, different tissues were isolated from 10 C57BL/6J mice and used to experimentally assess the methylation level of distal CpGs at the 5′ end of the candidate VM-IAPs using bisulfite pyrosequencing. IAPs with ragged methylation showed clear interindividual variation in methylation, while fully methylated IAPs did not (Figure 1B). Furthermore, these regions had consistent methylation levels across tissues within a single individual. Thus, within an individual animal, each VM-IAP exhibited its own level of methylation, and this level was evident in all somatic tissues of that individual, as described for A^{vy} and $Axin^{Fu}$ (Figure 1C). These findings confirm that ragged methylation can represent interindividual methylation variation, providing a framework for the second stage of the screen: the unbiased genome-wide identification of VM-IAPs.

Genome-wide Identification of VM-IAPs

To assess the full extent of VM-IAPs in the C57BL/6J genome, the ragged methylation pattern observed for our initial set of VM-IAPs was used to generate a genome-wide algorithm. The algorithm generated a value reflecting the methylation variation of each IAP in the C57BL/6J genome and was independent of expression of adjacent sequences. Computational variations at 68 IAP-LTRs (ranging from 5% to 64%) were experimentally verified to determine the accuracy of the algorithm and establish a threshold for true methylation variation (Figure S1A and Table S1). 25% methylation variation between the second-highest and second-lowest average methylation level between biological replicates was selected as the threshold for further analysis, as greater than 75% of randomly selected IAPs within this range showed more than 10% interindividual methylation variation upon experimental validation. This approach resulted in the identification of around 100 candidate VM-IAPs (Table S2).

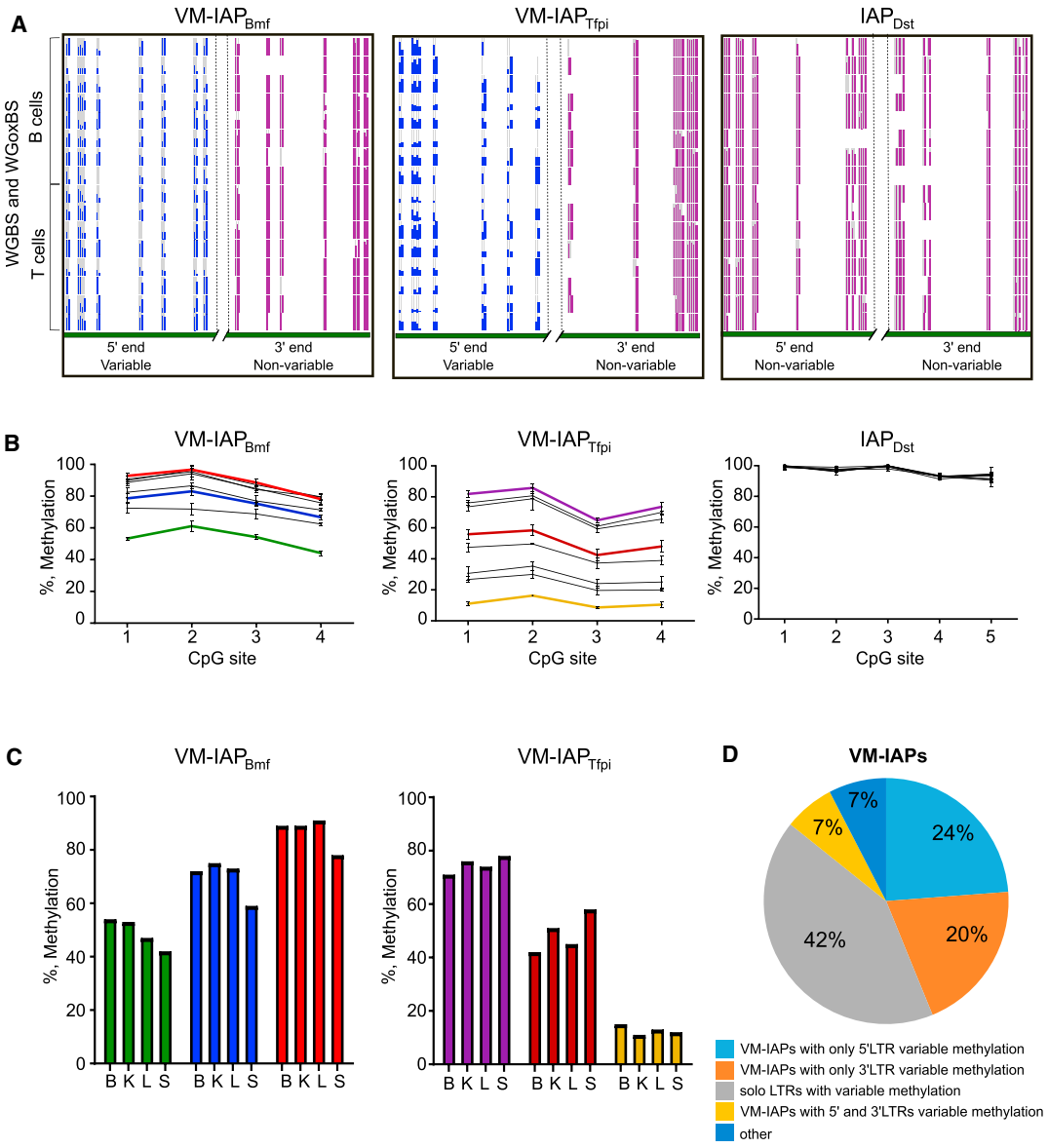


Figure 1. Identification of VM-IAPs

(A) WGBS and whole-genome oxidative bisulfite sequencing (WGoxBS) tracks of the distal regions of the 5′ and 3′ LTRs belonging to VM-IAP$_{Tfpi}$ (chr2:84505209-84510421), VM-IAP$_{Bmf}$ (chr2:118554765-118558375), and control hypermethylated IAP$_{Dst}$ (chr1:34347456-34351369). The poorly mapped central portion of the elements was removed due to its repetitive nature. Each vertical line represents one CpG, and each horizontal track represents one of 16 biological replicates. CpGs in variably methylated LTRs are highlighted in blue, illustrating ragged LTR methylation. CpGs in highly methylated LTRs are highlighted in purple.

(B) Bisulphite pyrosequencing validation of interindividual methylation levels in C57BL/6J kidney tissues (n = 8) at VM-IAP$_{Tfpi}$, VM-IAP$_{Bmf}$, and IAP$_{Dst}$. Each individual is represented by a single line. Sequenced CpGs are the most distal CpGs of the variably methylated LTRs. Colors are individual specific and correspond to those used in (C).

(C) Intraindividual methylation consistency at VM-IAP$_{Tfpi}$ and VM-IAP$_{Bmf}$ across brain ("B"), kidney ("K"), liver ("L"), and spleen ("S"). Individual-specific colors correspond to those in (E).

(D) Distribution of VM-IAPs according to methylation variation region. The number of VM-IAPs in each category is shown.

See also Figures S1 and S2 and Table S1.

We confirmed that the observed *in vivo* interindividual methylation variation was not a technical artifact resulting from tissue-type heterogeneity by repeating bisulfite pyrosequencing on pure populations of B cells from different mice and successfully validating interindividual methylation variation (Figure S1B). Given the well-established role of non-CpG methylation in transposable element silencing in plants (Stroud et al., 2014), we assessed VM-IAPs for non-CpG methylation. None was detected (Figure S1C).

VM-IAPs Are Evolutionarily Young Insertions with Locus-Specific Methylation States

A full-length intact IAP consists of 5′ and 3′ LTRs flanking retroviral genes (Falzon and Kuff, 1988; Mietz et al., 1987). Over evolutionary time, the structure of full-length IAPs is disrupted by deletions, point mutations, and recombination, eventually preventing retrotransposition activity (Stoye, 2001). The full range of IAP structures was represented in the identified set of candidate VM-IAPs, including full-length IAPs, truncated IAPs missing either their 5′ or 3′ LTR, solo LTRs, and truncated IAPs with no LTRs at all (Figure S2). Experimental validation of full-length VM-IAPs showed that 5′ LTRs differed from their corresponding 3′ LTRs in both their range and degree of methylation variation (Figures 1D and S2B). This indicates that 5′ and 3′ LTRs acquired their methylation variation independently of each other despite virtually identical genetic sequences.

The IAP nomenclature system used by the University of California, Santa Cruz (UCSC) Genome Browser classifies IAPs based on their LTR structural differences. This classification was used to investigate subtype enrichment of VM-IAPs. The majority of VM-IAPs had either IAPLTR1_Mm or IAPLTR2_Mm flanking LTRs (Figure 2A). Interestingly, IAPLTR1_Mm LTRs belonged to VM-IAPs with viral gene sequences while IAPLTR2_Mm elements were mainly solo LTR VM-IAPs. The IAPLTR1_Mm IAP subtype is considered to be the evolutionarily youngest IAP subtype (Qin et al., 2010). Furthermore, using data from the Mouse Genomes Project, we analyzed the presence or absence of VM-IAPs across 18 mouse strains and found that the majority of VM-IAPs were polymorphic insertions (Figure 2B).

Given the wide range of IAP structures and LTR subtypes, we investigated the possibility that the methylation state of VM-IAPs was exclusively determined by their genetic sequence. The IAPLTR1_Mm subtype was used to build a neighbor-joining tree using the sequences of all 780 IAPs of this LTR subtype in the C57BL/6J genome, including 27 VM-IAPs. Neighbor-joining tree analysis assesses sequence similarity and groups closely related IAPs together. Five distinct subtrees were identified containing VM-IAPs, with subtree 4 containing most of them (Figure S3A). This enrichment likely reflects their recent integration into the C57BL/6J genome and indicates that genetic sequence is at least partially involved in conferring methylation variation. Interestingly, VM-IAP$_{Slc15a2}$, IAP$_{Gpsm1}$ and IAP$_{Zak}$ are highly clustered, but only VM-IAP$_{Slc15a2}$ is a metastable epiallele, indicating that other factors, such as spatial organization, are likely at play (Figure S3B). We confirmed that sequence is not the sole determining factor for methylation level by comparing two VM-IAPs with 100% identical genetic sequences, showing they have different methylation levels within a given individual (Figure 2C).

We further explored co-variation of VM-IAP methylation levels within an individual by assessing the methylation state of six VM-IAPs in 33 different mice (Figure 2D). A normalized correlation matrix showed that the methylation level of each VM-IAP did not significantly correlate with that of other VM-IAPs within the same individual (Figure S3C). This indicates that the mechanism governing variable methylation likely acts in *cis* and argues against an overarching *trans*-mediated mechanism targeting all VM-IAPs within an individual in the same way.

VM-IAPs are Flanked by CTCF Binding Sites

Our results showed that other factors in addition to recent integration are involved in driving the methylation pattern observed at candidate VM-IAPs. We therefore asked whether the genomic location of VM-IAP insertion sites sets them apart from other non-variable IAPs. VM-IAPs are randomly distributed in the genome and do not cluster in specific topologically associating domains (TADs), nor are they enriched at TAD boundaries (data not shown). We found that approximately 70% of VM-IAPs are intergenic, and only two of them fall in UTRs. The remaining are intronic (Table S2A).

We next analyzed ENCODE chromatin immunoprecipitation (ChIP)-seq datasets to explore the epigenetic profiles of regions flanking VM-IAPs (ENCODE Project Consortium, 2012). It has been shown in embryonic stem cells (ESCs) that strain specific polymorphic IAP insertions are capable of spreading heterochromatic marks to flanking genomic DNA (Rebollo et al., 2011). However, no clear difference in H3K9me3 distribution was found between regions flanking VM-IAPs and non-variable IAPs (Figure S4A). Strikingly, the majority of VM-IAPs are bordered by CCCTC binding factor (CTCF) binding (Figures 3A). This enrichment was observed in datasets generated from different somatic tissues and from different developmental time points, suggesting stable maintenance of CTCF binding near VM-IAPs throughout development. To further investigate this enrichment, we produced a heatmap of the distance from the IAP border to the nearest CTCF peak across 14 ENCODE ChIP-seq datasets and found that CTCF was closer to VM-IAPs than non-variable IAPs across all datasets (Figure 3B). CTCF is a methylation-sensitive DNA binding protein that is crucial for both preimplantation and postimplantation stages (Phillips and Corces, 2009). CTCF-deficient oocytes cannot progress to the blastocyst stage following fertilization, and CTCF knockout embryos die before implantation (Wan et al., 2008; Moore et al., 2012). This developmental time point is in line with the proposed time point for the establishment of methylation at A^{vy} (Waterland et al., 2006; Blewitt et al., 2006). CTCF prefers unmethylated binding sites and has been shown to inhibit Dnmt1 activity to prevent methylation of its binding domain (Bell and Felsenfeld, 2000; Zampieri et al., 2012). It is an intriguing possibility that an interplay between IAP methylation and CTCF binding-site hypomethylation is involved in the establishment and/or maintenance of VM-IAPs.

VM-IAPs Can Interfere with Transcriptional Events

The methylation states of the IAPs integrated at the A^{vy} and *Axin*^{Fu} loci have been shown to have a direct impact on adjacent gene transcription and the phenotype of individual animals. In these cases, transcription initiates at the IAP LTR promoter, creating a chimeric transcript extending through the adjacent gene (Duhl et al., 1994; Vasicek et al., 1997). To explore the extent to which VM-IAPs can initiate transcription of endogenous genes, we analyzed *de novo* transcriptome assemblies generated from pure non-cycling populations of B and T cells (GEO: GSE94676). Only five VM-IAPs appeared to initiate transcripts overlapping annotated genes (Figure 4A). We investigated whether the expression level of these transcripts was related to the methylation level of the VM-IAPs and found statistically significant inverse correlations between gene expression and

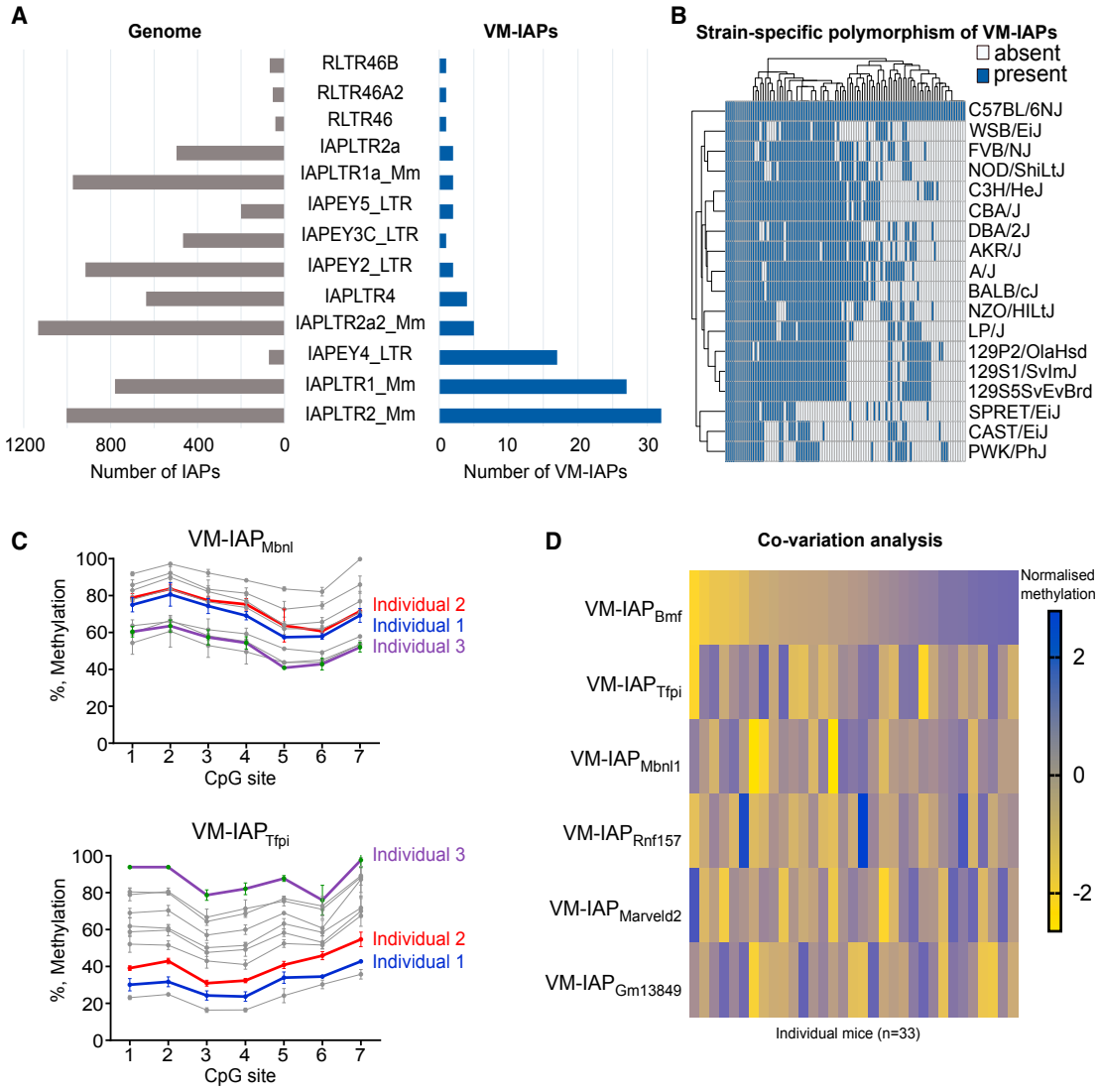


Figure 2. VM-IAPs Are Evolutionarily Young Insertions with Locus-Specific Methylation States

(A) Enrichment of IAP subtypes in VM-IAPs. The left side in gray represents the total number of IAPs in the genome of a particular LTR subtype. The right side in navy shows the number of VM-IAPs of a particular LTR subtype.

(B) Heatmap showing the presence or absence of VM-IAPs across 18 mouse strains, as determined from the Mouse Genomes Project (https://www.sanger.ac.uk/science/data/mouse-genomes-project). VM-IAPs are clustered by strain and presence of the IAP relative to C57BL/6J.

(C) Bisulphite pyrosequencing of VM-IAP$_{Tfpi}$ and VM-IAP$_{Mbnl1}$ in the same eight C57BL/6J mice. VM-IAP$_{Tfpi}$ and VM-IAP$_{Mbnl1}$ have identical sequences yet are differentially methylated within the same individual. Sequenced CpGs are the seven most distal ones of the 5′ LTRs. Specific individuals are color coded.

(D) Methylation levels at six VM-IAPs in 33 C57BL/6J mice are compared in a co-variation analysis. Methylation levels were normalized to a given VM-IAP's interindividual methylation range. Data show no relationship between VM-IAP methylation states within an individual.

See also Figure S3.

VM-IAP methylation levels, as observed for A^{vy} and $Axin^{Fu}$ (Figures 4B–4D). This finding reinforces the idea that metastable epialleles are capable of driving gene expression, yet reveals that promoter activity at these regions is an exception rather than the rule.

We next queried the number of VM-IAPs overlapping with transcripts in general and identified less than one-third of VM-IAPs with this property. Their insertion sites were enriched for the active histone marks H3K27ac, H3K9ac, and H3K4me3 (Figure S4B). No clear enrichment for H3K36me3 was observed (Figure S4B). It is therefore probable that the association of VM-IAPs with these transcripts is a product of insertion into transcriptionally active or "open" genomic regions. We found that half of the transcript-initiating or transcript-overlapping VM-IAPs were full-length IAPs with IAPLTR1_Mm LTRs, suggesting that evolutionarily young VM-IAPs are more likely to have a transcriptional influence (Figure 4A). Of note, CTCF binding at VM-IAP borders was independent of the presence or absence of overlapping

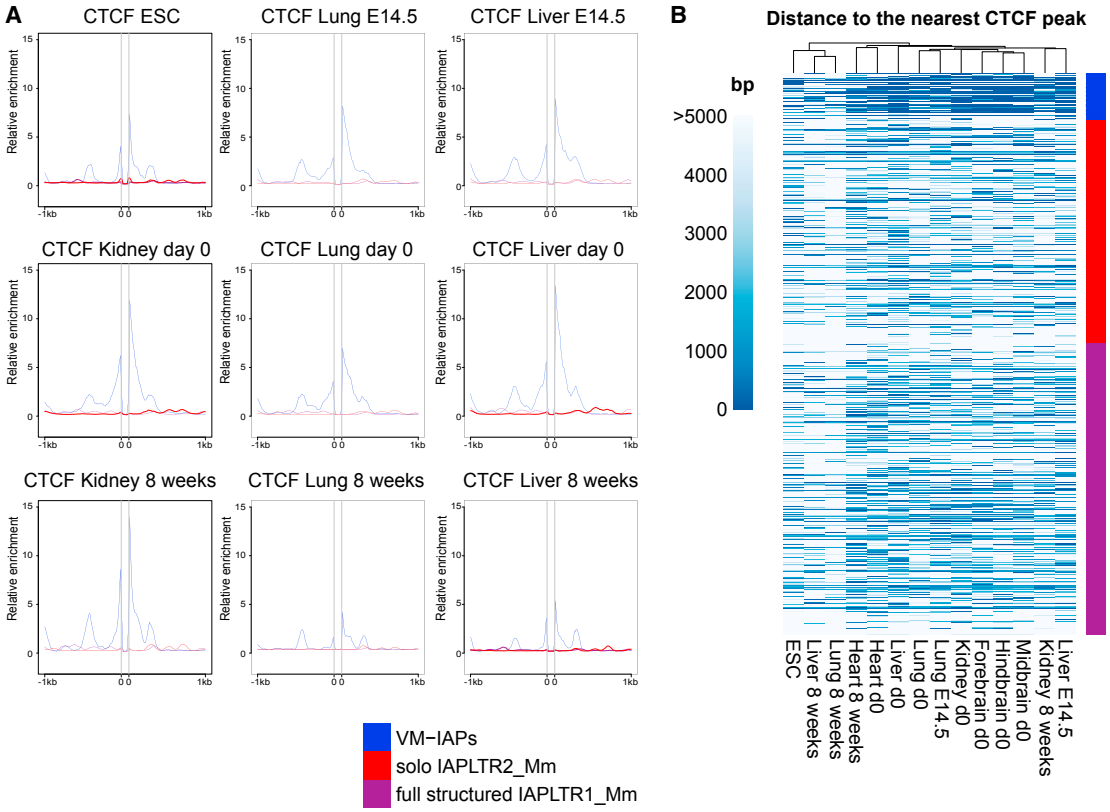


Figure 3. VM-IAPs Are Flanked by CTCF Binding Sites

(A) Relative CTCF enrichment profiles of VM-IAP flanking genomic regions in C57BL6/J ESCs, kidney, lung, and liver. Developmental stages E14.5, day 0, and 8 weeks are shown. Highly methylated full-length IAPs of the IAPLTR1_Mm subclass and solo LTRs of the IAPLTR2_Mm subclass serve as controls. All IAPs were fit to 100 bp, shown as the space between the two zeros. Zeros represent the start and end coordinates of IAPs. ChIP-seq datasets were downloaded from ENCODE (ENCODE Project Consortium, 2012).

(B) Heatmap of the distance from IAP border to the nearest CTCF peak across ChIP-seq ENCODE datasets. VM-IAPs, full structure IAPs of the IAPLTR1_Mm subclass, and solo LTRs of the IAPLTR2_Mm subclass are clustered and used in comparison.

See also Figure S4.

transcripts (Figure S4C), suggesting a potential functional role involving long-range interactions.

Methylation Variability Is Lost in the Male Germline and Re-established in the Next Generation

To evaluate heritability dynamics at VM-IAPs, we first assessed their methylation state in sperm. Nine VM-IAPs with wide methylation variation ranges were analyzed for LTR methylation levels in both somatic and mature sperm isolated from adult C57BL/6J male mice. As expected based on experimental validation, VM-IAPs showed interindividual methylation variation in the male somatic samples. In contrast, all VM-IAPs analyzed were fully methylated in sperm for all individuals, and no interindividual variation was observed (Figure 5A). Although repeat elements tend to be heavily methylated in sperm (Kobayashi et al., 2012), this is inconsistent with previous observations at A^{vy}, where partial methylation in sperm was observed to reflect the methylation status in somatic tissues despite the absence of heritability of phenotype from the sire (Rakyan et al., 2003).

We next sought to determine whether the methylation state of VM-IAPs is variable from one generation to the next. For the four VM-IAPs analyzed, the somatic methylation levels of pups born to highly or lowly methylated C57BL/6J mothers showed the full range of methylation variation observed in the previous generation (Figure 5B). This demonstrates that the methylation variability of VM-IAPs is faithfully reconstructed in the F1 generation after passage through the male and female germlines and that the methylation level of an individual does not influence its ability to produce offspring with the full methylation range associated with that VM-IAP.

To determine whether the inverse correlation between VM-IAP methylation and adjacent gene expression observed for a subset of VM-IAPs endures in the next generation, we assessed expression and methylation levels of VM-IAP$_{Eps8l1}$ and VM-IAP$_{Slc15a2}$ in maternal and offspring spleen tissues. We found that expression levels were variable among F1 littermates and inversely correlated with VM-IAP methylation levels (Figure S5A). Together, these findings indicate that the unique

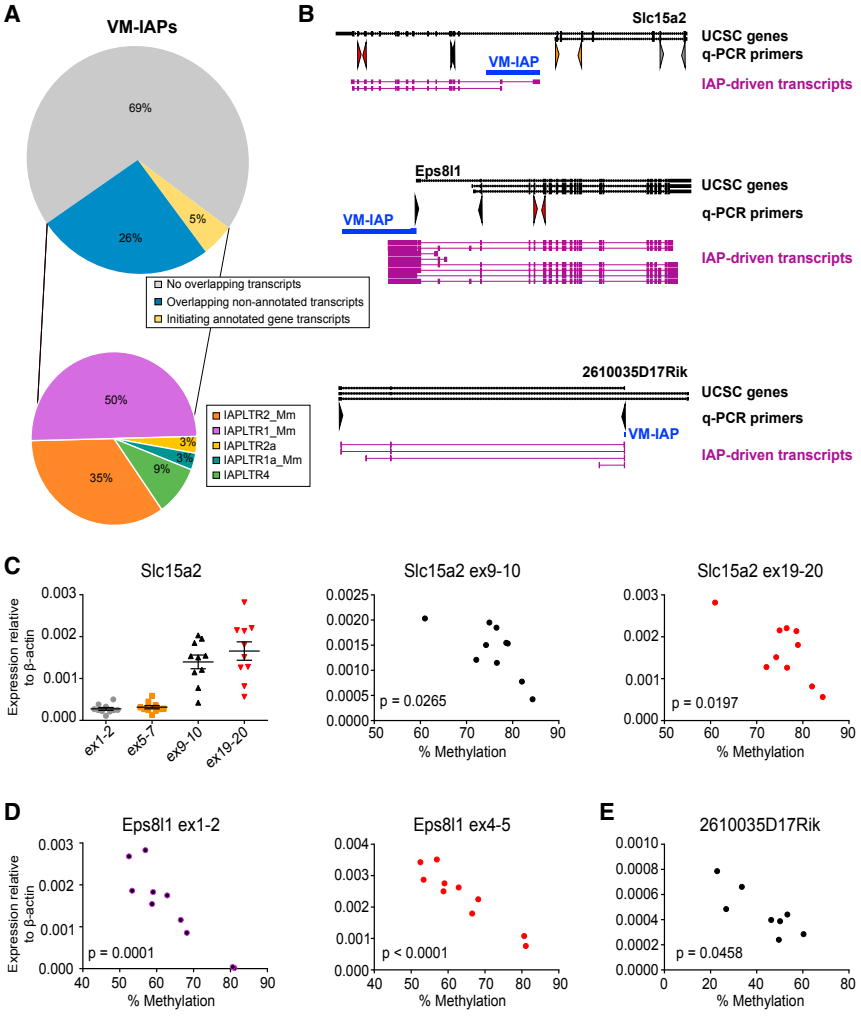


Figure 4. VM-IAPs Can Interfere with Tissue-Specific Transcriptional Events

(A) Pie charts of percentage of VM-IAPs overlapping *de novo* assembled transcripts and their distribution according to IAP-LTR subtype.

(B) VM-IAP-associated ectopic transcription of *Slc15a2*, *Eps8l1*, and *2610035D17Rik*. Intragenic VM-IAP$_{Slc15a2}$ drives expression of downstream *Slc15a2* exons. Intergenic VM-IAP$_{Eps8l1}$ drives ectopic expression of *Eps8l1*. Intragenic VM-IAP$_{2610035D17Rik}$ provides an alternative promoter for lincRNA *2610035D17Rik*. VM-IAPs are shown in blue and *de novo* assembled transcripts in purple. Transcripts extracted from UCSC are in black, and qPCR primers are depicted as arrows and color coded to correspond with (C), (D), and (E).

(C) Expression of *Slc15a2* downstream exons 9–10 and 19–20 (spleen) is inversely correlated with VM-IAP$_{Slc15a2}$ methylation (two-tailed Pearson). Upstream exons are not expressed. Expression was quantified by qPCR and shown relative to housekeeping gene *β-actin*. Each dot represents a different individual.

(D) Expression of *Eps8l1* exons 1–2 and 4–5 (brain) is inversely correlated with VM-IAP$_{Eps8l1}$ methylation (two-tailed Pearson).

(E) Expression of the VM-IAP$_{2610035D17Rik}$-driven transcript (spleen) is inversely correlated with VM-IAP$_{2610035D17Rik}$ methylation (two-tailed Pearson).

See also Figure S4.

methylation signature of each VM-IAP and its effect on transcription is re-established transgenerationally.

The A^{vy} locus is widely studied in large part because it exhibits epigenetic inheritance. In a C57BL/6J genetic background, the phenotype of the dam, but not the sire, influences the phenotypic distribution observed in the offspring (Morgan et al., 1999). In breeding-intense experiments, we investigated whether this was the case for six novel VM-IAPs. DNA methylation levels at the IAP LTR promoter were quantified in adult C57BL/6J breeding pairs and their offspring (Figures 6A–6F). We asked whether parental methylation level affected offspring methylation by building linear mixed-effects models (LMMs). Using LMMs allowed us to incorporate breeding pairs and litters as random effects, thereby controlling for the non-independence of siblings and littermates, respectively. We included maternal methylation level, paternal methylation level, and sex as potentially predictive fixed effects (Figure S5B).

For five out of the six regions tested, neither maternal nor paternal methylation level had a significant effect on offspring methylation levels (Figure 6G). For VM-IAP$_{Gm13849}$, the maternal methylation level, but not the paternal one, significantly affected offspring methylation levels ($p = 0.004$; $q = 0.036$; Figures 6G and S5B). This difference in heritability between parental lineages is consistent with the pattern observed for A^{vy} on a C57BL/6J background. Of note, while the maternal effect at this locus was significant, the R^2 was small, indicating that maternal methylation only accounts for a small fraction of the methylation variability observed in the next generation ($R^2 = 0.166$). Interestingly, when assessing whether the sex of an individual contributes to its methylation level, we found a highly significant sex effect at VM-IAP$_{Rnf157}$ whereby males are more likely to exhibit higher methylation levels than females ($p = 1.20e{-}5$; $q = 2.16e{-}4$; Figures 6G, S5B, and S5C). Weak evidence for a similar trend was observed for VM-IAP$_{Marveld2}$ and VM-IAP$_{Gm13849}$. This suggests there may be sex-linked modifiers of VM-IAPs.

Given that a single VM-IAP showed maternal heritability, we designed a smaller-scale experiment on VM-IAP$_{Gm13849}$ to validate this finding in a separate set of mice. We selected five highly methylated and five lowly methylated C57BL/6J females for breeding and subsequently assessed the VM-IAP$_{Gm13849}$ methylation levels of offspring from their first litter. We found that the methylation level of offspring born to highly methylated mothers was significantly different from that of offspring born to lowly methylated mothers, validating our previous result ($p = 0.0069$; $n = 5$ litters per group; Figure S5D). Together, our heritability studies indicate that inheritance of methylation levels is not a universal feature of VM-IAPs and instead show the remarkable reprogramming and faithful

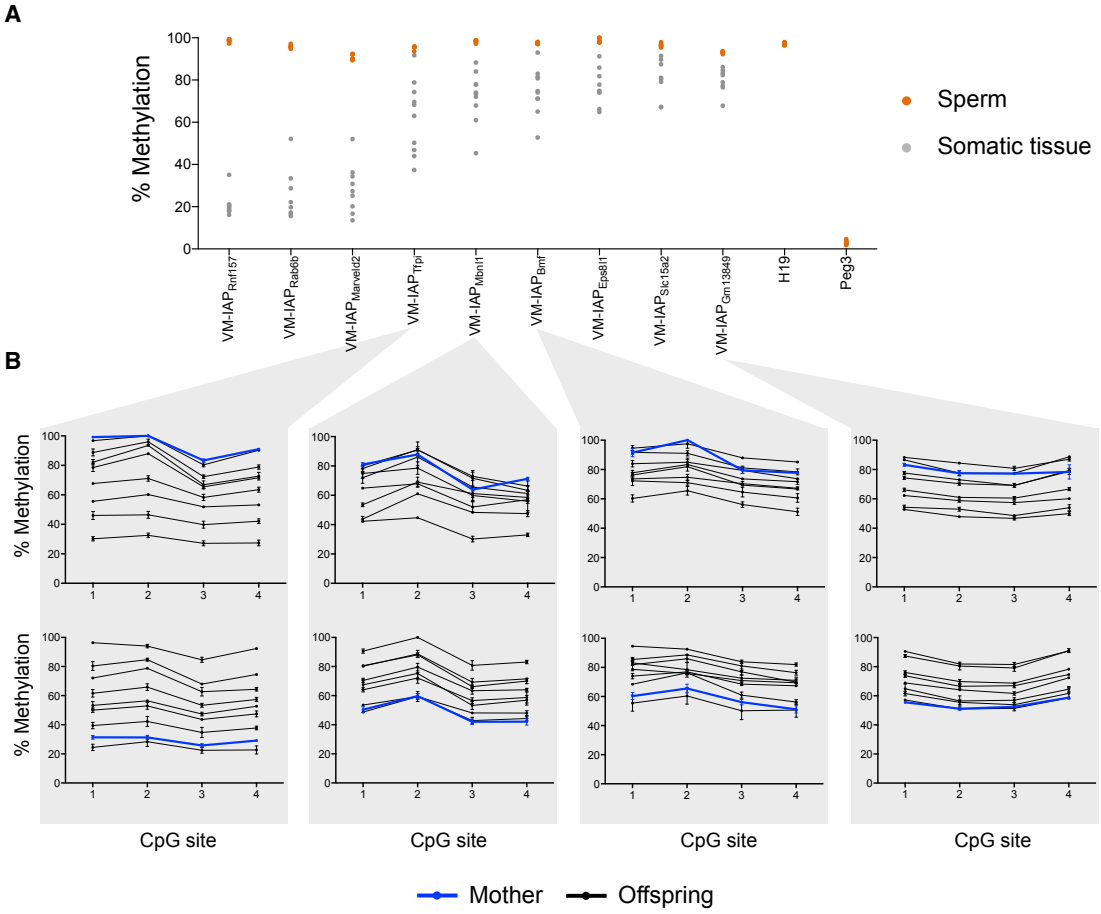


Figure 5. VM-IAPs Are Hypermethylated in the Male Germline and Reconstructed as Variable Loci in the Next Generation

(A) VM-IAPs are hypermethylated in sperm. Methylation levels at nine VM-IAPs in C57BL/6J sperm and corresponding somatic tissue are shown (n = 8–10 mice). Paternally expressed *Peg3* and maternally expressed *H19* serve as germ cell purity controls. Values shown are averages across the methylation level of the four distal CpG sites of the VM-IAP 5′ LTRs assessed by bisulphite pyrosequencing.

(B) Highly and lowly methylated mothers produce the full range of variably methylated offspring (n = 8 pups). Offspring and maternal methylation levels at VM-IAP$_{Tfpi}$, VM-IAP$_{Mbnl1}$, VM-IAP$_{Bmf}$, and VM-IAP$_{Gm13849}$ were assessed from ear samples by bisulphite pyrosequencing.

See also Figure S5.

re-establishment of VM-IAP variable states from one generation to the next regardless of parental methylation level.

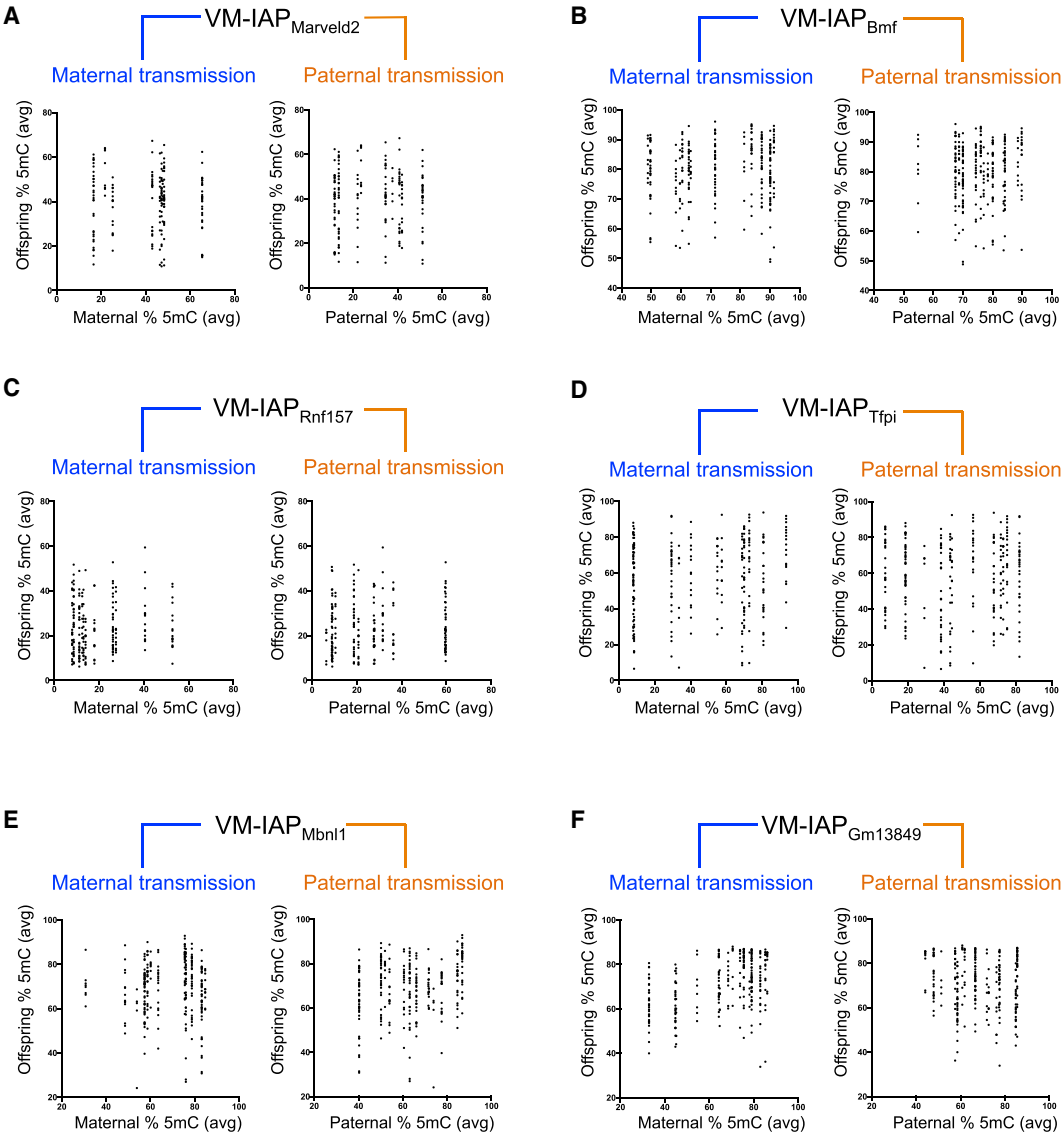
Non-IAP ERVs Can Exhibit Interindividual Methylation Variation

The same genome-wide screening strategy used to identify VM-IAPs was implemented to analyze other types of ERVs. Since the algorithm was based solely on the magnitude of methylation variation, it does not distinguish different ERV classes. 208 ERV1, 760 ERVK, and 174 ERVL candidates were identified as potential variably methylated ERVs (VM-ERVs) using the threshold levels developed in the initial model. 44 ERVs were randomly selected for experimental validation via bisulfite pyrosequencing. 13 of them validated as true VM-ERVs, showing more than 10% interindividual methylation variation (Table S3). Hence, although our WGBS-based screen identified a plethora of ERVs as VM-ERV candidates, less than 30% were true positives. This level of false positives might reflect the challenges of repetitive element alignment and/or differences in the CpG density at the LTR of different ERV subclasses influencing the application of the algorithm. Experimental validation and comparative analysis of these and the set of candidate VM-IAPs identified in this study will ascertain the full extent of epigenetic metastability at all LTR retrotransposons across the genome.

We computed the CpG density of different ERV-LTR subtypes in the genome and found that IAP LTRs indeed have the highest CpG density across all ERVs (Figure S6). It is therefore possible that CpG density might play a role in the properties of a metastable epiallele. This is in line with previous findings highlighting the particular susceptibility of IAPs to methylation-mediated regulation (Walsh et al., 1998).

DISCUSSION

A genome-wide screen identified multiple C57BL/6J VM-IAPs. Three of them (VM-IAP$_{Eps8l1}$, VM-IAP$_{Bmf}$, and VM-IAP$_{2610035D17Rik}$)



G

Linear mixed-effects models: effect sizes and statistical significance

Fixed effect	Maternal methylation			Paternal methylation			Sex			n
	R^{2*}	$p^{\dagger}$	$q^{\ddagger}$	R^2	p	q	R^2	p	q	
VM-IAP$_{Marveld2}$	0	0.849	0.849	0.001	0.776	0.847	0.026	0.022	0.113	183
VM-IAP$_{Bmf}$	0.021	0.240	0.540	0.001	0.800	0.847	0.003	0.310	0.569	278
VM-IAP$_{Rnf157}$	0.001	0.730	0.847	0.008	0.360	0.569	0.078	1.20e-5	**2.16e-4**	213
VM-IAP$_{Tfpi}$	0.020	0.171	0.513	0.001	0.757	0.847	0.006	0.210	0.540	264
VM-IAP$_{Mbnl1}$	0.012	0.323	0.569	0.068	0.048	0.173	0.001	0.547	0.757	265
VM-IAP$_{Gm13849}$	0.166	0.004	**0.036**	0.011	0.379	0.569	0.017	0.025	0.113	262

*R^2: semi-partial R^2 values for each fixed effect.
†p-values calculated using the Satterthwaite approximation for degrees of freedom.
‡q-values generated to correct for multiple testing using the Benjamini & Hochberg procedure (false discovery rate=0.05).

(legend on next page)

had previously been noted in a screen for the presence of H3K4me3 at retroelement promoters (Ekram et al., 2012). No H3K4me3 enrichment was found at lowly methylated A^{vy} individuals (Dolinoy et al., 2010)—this is consistent with our findings indicating a lack of H3K4me3 at VM-IAPs. A previous screen for metastable epialleles, encompassing all genomic regions but including only limited follow-up, identified 51 variably methylated ERV candidates. 22 overlapped with VM-IAPs identified in our study; the others were not detected in our screen (Oey et al., 2015). The previously identified metastable epiallele $IAP_{Cdk5rap1}$ (Druker et al., 2004) fell just below the stringent threshold of 25% used here. For this reason, in addition to the list of novel VM-IAPs, we have provided an additional set of IAPs picked up in our screen using the methylation variation observed at $IAP_{Cdk5rap1}$ as a threshold (Table S2, bottom section). However, this additional set contains several experimentally identified false positives and exhibits much narrower interindividual methylation ranges compared to the candidate VM-IAPs identified in our analysis.

Applying our screening strategy to other types of ERVs resulted in high false-positive rates upon experimental validation. This highlights the importance of rigorous experimental testing of candidates identified in studies on repeat elements and calls into question previous studies identifying non-IAP-derived metastable epialleles without conducting extensive validation. It remains challenging to develop bioinformatic approaches that distinguish epigenetically variable regions from poorly mapped ones. We nonetheless show that non-IAP ERVs are capable of exhibiting interindividual methylation variation and propose that the intriguing enrichment in IAP elements might reflect a requirement for high CpG density for the establishment of interindividual methylation variation.

We did not identify any specific genetic sequence features that explain acquisition of interindividual methylation variation at IAPs. In fact, very closely related VM-IAPs were found to have quite different ranges of methylation variation. In some cases, near-perfect sequence identity was observed between a VM-IAP and a non-variable IAP. However, we did find an enrichment for young classes of IAPs (IAPLTR1_Mm and IAPLTR2_Mm). It is therefore possible that VM-IAPs represent evolutionarily young IAPs in the process of becoming epigenetically silenced. In addition, VM-IAPs show an enrichment for strain-specific polymorphic IAPs, suggesting that recent integration into the C57BL/6J genome may contribute to their variably methylated states. Further analysis of the relationship between the acquisition of variable methylation and the strain-specific placement of such retrotransposons will contribute to our understanding of the mechanism through which they escape the fate of their fully methylated counterparts.

Unlike A^{vy}, the majority of VM-IAPs described in this study do not initiate transcription events, indicating that these elements are not generally functioning as heterologous promoters. We found, however, that transcript expression that is initiated at VM-IAPs is low, and we detected no ontology enrichment for those genes adjacent to VM-IAPs.

Most compelling, however, is the finding that VM-IAPs are enriched for CTCF binding in their flanking regions. Since CTCF is a methylation-sensitive DNA binding protein preferring unmethylated DNA (Phillips and Corces, 2009), this association suggests a functional antagonism between methylation of an IAP and the maintenance of an unmethylated state at the CTCF binding site, leading to the early acquisition of a stochastically methylated state. Extensive further experiments will be required to decipher the potential relationship between CTCF binding and VM-IAPs. It has been proposed that VM-IAPs may have a selective advantage, conferring stochastic fitness and enhanced evolution (Branciamore et al., 2014, 2015). Taking into consideration a potential role for VM-IAPs in influencing long-range *cis*-acting interactions and the interstrain differences in the absence or presence of some of these elements, this idea can now be tested experimentally through comparative functional analyses between individuals and between different genetic backgrounds.

The hypermethylation observed in sperm at VM-IAPs, reconstructed into a variably methylated state after fertilization, suggests reprogramming upon paternal inheritance. Given that the methylation state of the A^{vy} and $Axin^{Fu}$ alleles in sperm has been reported to reflect the methylation state in somatic tissues of the same individual (Rakyan et al., 2003; Blewitt et al., 2006), one might have expected the VM-IAPs identified in this study to behave in a similar manner. It is possible, however, that A^{vy} and $Axin^{Fu}$ alleles are unusual compared to other metastable epialleles. Alternatively, this inconsistency may reflect the fact that they arose as insertional mutations in non-C57BL/6J mouse strains but have been maintained on a C57BL/6J background for the experiments in question. Hence, these loci, being the only non-C57BL/6J segments of DNA in an otherwise C57BL/6J genome, may be more refractory to the C57BL/6J sperm methylation machinery. Alternatively, the difference may be technical, reflecting the different methods used to quantify methylation. Despite the acquisition of complete methylation occurring in the male germline, we show that VM-IAPs are re-established

Figure 6. Inheritance Analysis of VM-IAPs

(A–F) Methylation levels at VM-IAPs were quantified from ear samples taken from C57BL/6J breeding pairs and their offspring. Offspring methylation level is plotted against maternal and paternal methylation level for $VM\text{-}IAP_{Marveld2}$ (A), $VM\text{-}IAP_{Bmf}$ (B), $VM\text{-}IAP_{Rnf157}$ (C), $VM\text{-}IAP_{Tfpi}$ (D), $VM\text{-}IAP_{Mbnl1}$ (E), and $VM\text{-}IAP_{Gm13849}$ (F). Methylation levels represent averages across the first four distal CpG sites of the VM-IAP 5′ LTRs.

(G) Statistical output from LMMs showing a significant maternal effect on offspring methylation at $VM\text{-}IAP_{Gm13849}$ as well as a significant sex effect on $VM\text{-}IAP_{Rnf157}$ methylation levels. Analyses were carried out using the lmer() function from the lme4 R package with maternal methylation level, paternal methylation level, and sex treated as fixed effects. Breeding pair and litter were treated as random effects. Raw p values were generated using the Satterthwaite approximation for degrees of freedom. Adjusted p values (q) were generated using the Benjamini and Hochberg correction to account for multiple testing. Semi-partial R^2 values, representing the effect size for each fixed effect, were calculated using the r2beta() function from the r2glmm R package. Sample sizes are shown.

See also Figure S5.

as variable loci in the next generation, indicating faithful reconstruction of their variable state from one generation to the next.

Only one of the VM-IAPs we examined showed evidence of a maternal methylation-level memory reminiscent of previously described A^{vy} inheritance dynamics in a C57BL/6J genetic background. For this lone VM-IAP, the highly quantitative nature of the methylation analysis indicated a small effect size. Our findings raise questions about the generalizability of non-genetic inheritance at metastable epialleles and suggest that variable methylation can be reprogrammed and reconstructed across generations in the absence of a memory of parental state by a process that may depend on the genetic context of the variably modified locus.

Importantly, our inheritance analysis highlights how consistent the range of interindividual methylation variation is at each VM-IAP regardless of parental methylation state. The mechanism by which the variable state of these unique elements is reprogrammed and precisely re-established in the next generation remains to be elucidated. Interestingly, the bordering regions of previously identified putative human metastable epialleles are enriched for ERVs and long interspersed nuclear elements (LINEs), suggesting that the link between repeat elements and epigenetic metastability may be conserved (Silver et al., 2015). The findings reported here establish a repertoire of murine loci to study mechanisms of non-genetic inheritance, the influence of the repeat genome on phenotype, and the epigenetic impact of normal and compromised environmental contexts.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - MICE
- METHOD DETAILS
  - Tissue and B cell collection
  - DNA/RNA extraction and bisulfite conversion
  - DNA methylation analysis
  - Expression analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Biased screen of polymorphic IAPs
  - Genome-wide screen
  - Strain-specific polymorphism analysis
  - Neighbor-joining tree analysis
  - Co-variation analysis
  - Generation of enrichment profiles
  - Transcriptomic analysis
  - Inheritance analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and six tables and can be found with this article online at https://doi.org/10.1016/j.cell.2018.09.043.

ACKNOWLEDGMENTS

Research was funded by BLUEPRINT (HEALTH-F5-2011-282510) to A.C.F.-S. and D.A. and by the Wellcome Trust, United Kingdom (WT095606RR) and Medical Research Council, United Kingdom (MR/J001597/1) to A.C.F.-S. We are grateful for PhD studentships from the Darwin Trust to A.K.; from the Cambridge Trust, Downing College, and Pomona College to T.M.B.; and from the BBSRC to J.G. and R.G. We are very grateful to Tony Fulford for his advice on the statistical analysis of inheritance data. We also thank Rahia Mashoodh, Dionne Gray, Jennifer Corish, and other members of the Ferguson-Smith lab for technical support and helpful discussions, Ali Robinson for animal support during the breeding experiments, and Antara Mujamdar and James Lee (SUNY Buffalo) for technical assistance.

AUTHOR CONTRIBUTIONS

A.K. conducted the screen and VM-IAP characterization. T.M.B. carried out the germline and inheritance experiments. M.K.S.-H. performed the experimental work, and N.W. performed the bioinformatics associated with the BLUEPRINT datasets. D.A. and A.C.F.-S. designed the BLUEPRINT mouse model work package. J.G. and R.G. contributed to the analysis of transcriptomes, and E.P. contributed to the analysis of their initiation and termination. S.A. provided data on variable methylation validation. A.C.F.-S. conceived and designed the project. A.K., T.M.B., and A.C.F.-S. drafted the manuscript. All authors commented on the manuscript.

DECLARATION OF INTERESTS

A.C.F.-S. is a consultant for Sanofi.

Received: January 29, 2018
Revised: July 19, 2018
Accepted: September 19, 2018
Published: October 25, 2018; corrected online: November 7, 2018

REFERENCES

Adams, D., Altucci, L., Antonarakis, S.E., Ballesteros, J., Beck, S., Bird, A., Bock, C., Boehm, B., Campo, E., Caricasole, A., et al. (2012). BLUEPRINT to decode the epigenetic signature written in blood. Nat. Biotech. *30*, 224–226.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. *67*, 1–48.

Becker, C., and Weigel, D. (2012). Epigenetic variation: origin and transgenerational inheritance. Curr. Opin. Plant Biol. *15*, 562–567.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482–485.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B *57*, 289–300.

Blewitt, M.E., Vickaryous, N.K., Paldi, A., Koseki, H., and Whitelaw, E. (2006). Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. PLoS Genet. *2*, e49.

Branciamore, S., Rodin, A.S., Riggs, A.D., and Rodin, S.N. (2014). Enhanced evolution by stochastically variable modification of epigenetic marks in the early embryo. Proc. Natl. Acad. Sci. USA *111*, 6353–6358.

Branciamore, S., Rodin, A.S., Gogoshin, G., and Riggs, A.D. (2015). Epigenetics and Evolution: Transposons and the Stochastic Epigenetic Modification Model. AIMS Genetics *2*, 148–162.

Cubas, P., Vincent, C., and Coen, E. (1999). An epigenetic mutation responsible for natural variation in floral symmetry. Nature *401*, 157–161.

Daxinger, L., Harten, S.K., Oey, H., Epp, T., Isbel, L., Huang, E., Whitelaw, N., Apedaile, A., Sorolla, A., Yong, J., et al. (2013). An ENU mutagenesis screen identifies novel and known genes involved in epigenetic processes in the mouse. Genome Biol. *14*, R96.

Dickies, M.M. (1962). A new viable yellow mutation in the house mouse. J. Hered. *53*, 84–86.

Dolinoy, D.C., Weidman, J.R., Waterland, R.A., and Jirtle, R.L. (2006). Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. Environ. Health Perspect. *114*, 567–572.

Dolinoy, D.C., Huang, D., and Jirtle, R.L. (2007). Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. Proc. Natl. Acad. Sci. USA *104*, 13056–13061.

Dolinoy, D.C., Weinhouse, C., Jones, T.R., Rozek, L.S., and Jirtle, R.L. (2010). Variable histone modifications at the A(vy) metastable epiallele. Epigenetics *5*, 637–644.

Druker, R., Bruxner, T.J., Lehrbach, N.J., and Whitelaw, E. (2004). Complex patterns of transcription at the insertion site of a retrotransposon in the mouse. Nucleic Acids Res. *32*, 5800–5808.

Duhl, D.M., Vrieling, H., Miller, K.A., Wolff, G.L., and Barsh, G.S. (1994). Neomorphic agouti mutations in obese yellow mice. Nat. Genet. *8*, 59–65.

Ekram, M.B., Kang, K., Kim, H., and Kim, J. (2012). Retrotransposons as a major source of epigenetic variations in the mammalian genome. Epigenetics *7*, 370–382.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Falzon, M., and Kuff, E.L. (1988). Multiple protein-binding sites in an intracisternal A particle long terminal repeat. J. Virol. *62*, 4070–4077.

Faulk, C., Barks, A., and Dolinoy, D.C. (2013). Phylogenetic and DNA methylation analysis reveal novel regions of variable methylation in the mouse IAP class of transposons. BMC Genomics *14*, 48.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., et al. (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. *44* (D1), D279–D285.

Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., et al. (2006). The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. *34*, D590–D598.

Jaeger, B.C., Edwards, L.J., Das, K., and Sen, P.K. (2017). An R^2 statistic for fixed effects in the generalized linear mixed model. J. Appl. Stat. *44*, 1086–1105.

Kaminen-Ahola, N., Ahola, A., Maga, M., Mallitt, K.A., Fahey, P., Cox, T.C., Whitelaw, E., and Chong, S. (2010). Maternal ethanol consumption alters the epigenotype and the phenotype of offspring in a mouse model. PLoS Genet. *6*, e1000811.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, D493–D496.

Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., et al. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. Nature *477*, 289–294.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.

Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y., et al. (2012). Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. PLoS Genet. *8*, e1002440.

Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. J. Stat. Softw. *82*, 1–26.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Michaud, E.J., van Vugt, M.J., Bultman, S.J., Sweet, H.O., Davisson, M.T., and Woychik, R.P. (1994). Differential expression of a new dominant agouti allele (A^{iapy}) is correlated with methylation state and is influenced by parental lineage. Genes Dev. *8*, 1463–1472.

Mietz, J.A., Grossman, Z., Lueders, K.K., and Kuff, E.L. (1987). Nucleotide sequence of a complete mouse intracisternal A-particle genome: relationship to known aspects of particle assembly and function. J. Virol. *61*, 3020–3029.

Miska, E.A., and Ferguson-Smith, A.C. (2016). Transgenerational inheritance: Models and mechanisms of non-DNA sequence-based inheritance. Science *354*, 59–63.

Moore, J.M., Rabaia, N.A., Smith, L.E., Fagerlie, S., Gurley, K., Loukinov, D., Disteche, C.M., Collins, S.J., Kemp, C.J., Lobanenkov, V.V., and Filippova, G.N. (2012). Loss of maternal CTCF is associated with peri-implantation lethality of Ctcf null embryos. PLoS ONE *7*, e34915.

Morgan, D.K., and Whitelaw, E. (2008). The case for transgenerational epigenetic inheritance in humans. Mamm. Genome *19*, 394–397.

Morgan, H.D., Sutherland, H.G., Martin, D.I., and Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. Nat. Genet. *23*, 314–318.

Nellåker, C., Keane, T.M., Yalcin, B., Wong, K., Agam, A., Belgard, T.G., Flint, J., Adams, D.J., Frankel, W.N., and Ponting, C.P. (2012). The genomic landscape shaped by selection on transposable elements across 18 mouse strains. Genome Biol. *13*, R45.

Oey, H., Isbel, L., Hickey, P., Ebaid, B., and Whitelaw, E. (2015). Genetic and epigenetic variation among inbred mouse littermates: identification of inter-individual differentially methylated regions. Epigenetics Chromatin *8*, 54.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295.

Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 1194–1211.

Qin, C., Wang, Z., Shang, J., Bekkari, K., Liu, R., Pacchione, S., McNulty, K.A., Ng, A., Barnum, J.E., and Storer, R.D. (2010). Intracisternal A particle genes: Distribution in the mouse genome, active subtypes, and potential roles as species-specific mediators of susceptibility to cancer. Mol. Carcinog. *49*, 54–67.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

R Development Core Team (2016). R: A language and environment for statistical computing (R Foundation for Statistical Computing).

Rakyan, V.K., Blewitt, M.E., Druker, R., Preis, J.I., and Whitelaw, E. (2002). Metastable epialleles in mammals. Trends Genet. *18*, 348–351.

Rakyan, V.K., Chong, S., Champ, M.E., Cuthbert, P.C., Morgan, H.D., Luu, K.V., and Whitelaw, E. (2003). Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. Proc. Natl. Acad. Sci. USA *100*, 2538–2543.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, (Web Server issue), W187-91.

Rebollo, R., Karimi, M.M., Bilenky, M., Gagnier, L., Miceli-Royer, K., Zhang, Y., Goyal, P., Keane, T.M., Jones, S., Hirst, M., et al. (2011). Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. PLoS Genet. *7*, e1002301.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Sharma, U., Conine, C.C., Shea, J.M., Boskovic, A., Derr, A.G., Bing, X.Y., Belleannee, C., Kucukural, A., Serra, R.W., Sun, F., et al. (2016). Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. Science *351*, 391–396.

Silver, M.J., Kessler, N.J., Hennig, B.J., Dominguez-Salas, P., Laritsky, E., Baker, M.S., Coarfa, C., Hernandez-Vargas, H., Castelino, J.M., Routledge, M.N., et al. (2015). Independent genomewide screens identify the tumor

suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment. Genome Biol. *16*, 118.

Smit, A.F.A., Hubley, R., and Green, P. (2015). RepeatMasker Open-4.0.

Stoye, J.P. (2001). Endogenous retroviruses: still active after all these years? Curr. Biol. *11*, R914–R916.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat. Struct. Mol. Biol. *21*, 64–72.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3–new capabilities and interfaces. Nucleic Acids Res. *40*, e115.

Vasicek, T.J., Zeng, L., Guan, X.J., Zhang, T., Costantini, F., and Tilghman, S.M. (1997). Two dominant mutations in the mouse fused gene are the result of transposon insertions. Genetics *147*, 777–786.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat. Genet. *20*, 116–117.

Wan, L.B., Pan, H., Hannenhalli, S., Cheng, Y., Ma, J., Fedoriw, A., Lobanenkov, V., Latham, K.E., Schultz, R.M., and Bartolomei, M.S. (2008). Maternal depletion of CTCF reveals multiple functions during oocyte and preimplantation embryo development. Development *135*, 2729–2739.

Waterland, R.A., Dolinoy, D.C., Lin, J.R., Smith, C.A., Shi, X., and Tahiliani, K.G. (2006). Maternal methyl supplements increase offspring DNA methylation at Axin Fused. Genesis *44*, 401–406.

Weinhouse, C., Anderson, O.S., Jones, T.R., Kim, J., Liberman, S.A., Nahar, M.S., Rozek, L.S., Jirtle, R.L., and Dolinoy, D.C. (2011). An expression microarray approach for the identification of metastable epialleles in the mouse genome. Epigenetics *6*, 1105–1113.

Wolff, G.L., Kodell, R.L., Moore, S.R., and Cooney, C.A. (1998). Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. FASEB J. *12*, 949–957.

Yalcin, B., Wong, K., Agam, A., Goodson, M., Keane, T.M., Gan, X., Nellåker, C., Goodstadt, L., Nicod, J., Bhomra, A., et al. (2011). Sequence-based characterization of structural variation in the mouse genome. Nature *477*, 326–329.

Zampieri, M., Guastafierro, T., Calabrese, R., Ciccarone, F., Bacalini, M.G., Reale, A., Perilli, M., Passananti, C., and Caiafa, P. (2012). ADP-ribose polymers localized on Ctcf-Parp1-Dnmt1 complex prevent methylation of Ctcf target sites. Biochem. J. *441*, 645–652.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., et al. (2018). Ensembl 2018. Nucleic Acids Res. *46* (D1), D754–D761.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
B Cell Isolation Kit	Miltenyi Biotec	Cat#130-090-862
AllPrep DNA/RNA Mini Kit	QIAGEN	Cat#80204
RNase-Free DNase Set	QIAGEN	Cat#79254
PyroMark Gold Q96 Reagents	QIAGEN	Cat#972804
Imprint DNA Modification Kit	Sigma	Cat#MOD50
RevertAid H Minus First Strand cDNA Synthesis Kit	Thermo Fisher Scientific	Cat#K1631
LightCycler 480 SYBR Green I Master	Roche	Cat#04707516001
HotStarTaq DNA Polymerase	QIAGEN	Cat#203203
Deposited Data		
WGBS, WGoxBS, and RNA-seq datasets	This paper	GEO: GSE94676
Experimental Models: Organisms/Strains		
Mouse: C57BL/6J	The Jackson Laboratory	JAX: 000664
Oligonucleotides		
qPCR primers, see Table S4	This paper	N/A
PCR and pyrosequencing primers, see Table S5	This paper	N/A
Software and Algorithms		
Prism 7 for Mac OS X	GraphPad Software	https://www.graphpad.com/scientific-software/prism/
Geneious 9.0.5	Kearse et al., 2012	https://www.geneious.com/
PyroMark Assay Design SW 2.0	QIAGEN	Cat#9019077
Primer3	Untergasser et al., 2012	http://primer3.ut.ee/
liftOver	Hinrichs et al., 2006	http://genome.ucsc.edu/
BEDTools 2.25.0	Quinlan and Hall, 2010	https://bedtools.readthedocs.io/en/latest/index.html
DESeq2 1.3.52	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
edgeR 3.5.27	Robinson et al., 2010	https://bioconductor.org/packages/release/bioc/html/edgeR.html
RepeatMasker	Smit et al., 2015	http://www.repeatmasker.org/
Mouse Genomes Project	Nellåker et al., 2012 Keane et al., 2011 Yalcin et al., 2011	https://www.sanger.ac.uk/sanger/Mouse_SnpViewer/rel-1505
ENCODE	ENCODE Consortium	https://www.encodeproject.org/
Galaxy deepTools	Ramírez et al., 2014	https://usegalaxy.org
StringTie 1.3.3	Pertea et al., 2015	https://ccb.jhu.edu/software/stringtie/#pub
R	R Development Core Team, 2016	https://www.r-project.org/
Pfam	Finn et al., 2016	https://pfam.xfam.org/
Ensembl annotation	Zerbino et al., 2018	http://www.ensembl.org/index.html?redirect=no
Custom R code for data analysis (Github)	This paper	https://github.com/AFS-lab/Kazachenka-Bertozzi-et-al-2018
Other		
Streptavidin Sepharose High Performance	GE Healthcare	Cat#GE17-5113-01

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Anne C. Ferguson-Smith (afsmith@mole.bio.cam.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

MICE

All mouse work was carried out in accordance with UK government Home Office licensing procedures (HO project license number: PC9886123). All experiments used C57BL/6J mice of both sexes. The methylation validation and expression experiments were performed on 8-10 week old mice. For the inheritance studies, mice were set up for breeding at 8 weeks of age and the F1 methylation level was assessed in ear notches from 10-12 day old pups. All mice were fed a standard chow diet *ad libitum* and housed in controlled temperature, humidity, and light-dark cycle (12h) conditions.

METHOD DETAILS

Tissue and B cell collection

Following dissection, somatic C57BL/6J tissues were snap frozen in liquid nitrogen and manually pulverized. B cells were isolated from fresh splenic tissues using the B Cell Isolation Kit (Miltenyi Biotec). Sperm collection and purification from cauda epididymis was done as described in Sharma et al. (2016).

DNA/RNA extraction and bisulfite conversion

30 ug of tissue (brain, kidney, liver, and spleen) was used for simultaneous purification of genomic DNA and total RNA using the AllPrep DNA/RNA Mini Kit (QIAGEN). During purification, RNA was treated with DNaseI using the RNase-Free DNase Set (QIAGEN). Ear notch DNA was purified using a standard phenol-chloroform extraction protocol. DNA was bisulfite treated using the two-step protocol of the Imprint DNA Modification Kit (Sigma).

DNA methylation analysis

Methylation quantification was carried out by pyrosequencing. Assays were designed using PyroMark Assay Design SW 2.0 (QIAGEN). Primers are provided in Table S4. Regions of interest were amplified from bisulfite converted DNA via PCR using biotinylated reverse primers and HotStarTaq DNA Polymerase (QIAGEN). The annealing temperature for PCR primers was optimized by gradient PCR. PCR conditions: 1) 95°C – 5 min; 2) 94°C – 30 s, optimized t°C – 30 s, 72°C – 55 s, 40 cycles; 3) 72°C – 5 min. PCR products were shaken at 1,400 rpm with Streptavidin Sepharose High Performance beads (GE healthcare) dissolved in binding buffer (10mM Tris-HCL pH7.6, 2M NaCl, 1mM EDTA, 0.1% Tween-20) for 20 min. The biotinylated strand was purified using the PyroMark Q96 Vacuum Workstation (QIAGEN). Sequencing primers were annealed to the template in annealing buffer (20mM Tris-acetate pH7.6, 2M magnesium acetate) at 85°C for 3 min. Sequencing was carried out on the PyroMark Q96 MD pyrosequencer (QIAGEN) using PyroMark Gold Q96 Reagents (QIAGEN).

Expression analysis

cDNA was synthesized using the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). Q-PCR primers were designed using Primer3 software (Untergasser et al., 2012) and are listed in Table S4. cDNA was amplified using the LightCycler 480 SYBR Green I Master mix and LightCycler 480 Instrument (Roche). PCR conditions: 1) 95°C – 5 min; 2) 95°C – 10 s, 60°C – 10 s, 72°C – 10 s, 45 cycles; 3) 95°C - 5 s, 65°C - 1 min, 97°C - continuous ; 4) 40°C – 30 s. Relative cDNA abundance was calculated using the ΔCT method and normalized to housekeeping gene *β-actin*. The significance of correlations between expression and methylation levels was assessed by computing Pearson correlation coefficients followed by two-tailed p values in GraphPad Prism.

QUANTIFICATION AND STATISTICAL ANALYSIS

Biased screen of polymorphic IAPs

Genomic coordinates of C57BL/6J-specific IAPs that are absent from the CAST/Eij genome were extracted from a published list of polymorphic ERVs (Nellåker et al., 2012). IAP coordinates were converted from the mm9 to the mm10 mouse genome assembly using liftOver (Hinrichs et al., 2006) and assigned to the nearest protein-coding gene from the Ensembl gene database (GRCm38) using Bedtools (Quinlan and Hall, 2010). DESeq2 and edgeR were used to identify differentially expressed (DE) genes for B and T cell samples. Due to the absence of strain-specific annotation of ncRNAs, they were removed from the analysis. Significant hits from both programs were used to compile the list of DE genes between the two strains. DE genes were overlapped with genes containing a polymorphic IAP insertion or having one nearby. The overlapping genes provided a list of candidate IAPs for visual assessment of methylation levels.

Genome-wide screen

The genome-wide screen for VM-IAPs used WGBS and WGoxBS datasets generated from B and T cells (16 datasets in total; accession number: GSE94676). WGBS and WGoxBS datasets were treated as biological replicates, as the WGoxBS protocol recognizes both methylated and hydroxymethylated DNA. The 5mC:5hmC ratio in the WGoxBS datasets used was 1:0.015. Each biological

replicate consisted of pooled B or T cells for 4-5 individuals. Bedgraph files were used to extract methylation levels of IAP CpGs. Methylation of a CpG site is represented by two values, reflecting methylation levels of the sense and antisense strands. The average of 16 methylation values representing 8 distal CpGs from the 5′ or 3′ end of IAPs was calculated for each biological replicate to estimate a given IAP's methylation level. To determine the magnitude of methylation variation at an IAP across the 16 biological replicates, the average methylation levels were sorted and the difference between the second highest and the second lowest values was used as a computational score for methylation variation (Figure S1). The analysis for 5′ and 3′ ends was done separately and subsequently overlapped. The threshold of variation used as a cut-off for the final list of VM-IAPs was determined by experimental assessment of the methylation variation of a subset of IAPs exhibiting a range of computational scores (Table S5). Differentially methylated regions between B and T cells were excluded from the final list of VM-IAPs since these constituted cell type-specific DMRs and hence did not fulfill the criteria of methylation consistency between tissues. An IAP was considered to be a DMR if its 8 highest and/or 8 lowest average methylation levels came from only one of the two cell types. The same process was carried out for sex-specific DMRs, but none were found.

Strain-specific polymorphism analysis

A catalog of structural variants across 18 inbred mouse strains generated for the Mouse Genomes Project (https://www.sanger.ac.uk/sanger/Mouse_SnpViewer/rel-1505) was used to quantify polymorphism of VM-IAP candidates (Keane et al., 2011; Yalcin et al., 2011).

Neighbor-joining tree analysis

The neighbor-joining tree of IAPs of the IAPLTR1_Mm subtype was built with Geneious 9.0.5 software using default parameters (Kearse et al., 2012; Faulk et al., 2013). IAP sequences were downloaded from the UCSC Table Browser and "+" strand sequences were used for antisense IAPs.

Co-variation analysis

Methylation levels at six VM-IAPs in 33 C57BL/6J mice were normalized to a given VM-IAP's inter-individual methylation range. A normalized correlation matrix showing Pearson correlation coefficients as well as p values for the correlation of each VM-IAP pair was generated using GraphPad Prism. A Bonferroni-adjusted α value of 0.008 was used.

Generation of enrichment profiles

Histone modification and CTCF binding profiles were constructed using publicly available ENCODE datasets. Accession numbers for all ENCODE datasets used can be found in Table S6. Signal p value bigwig files were downloaded and analyzed using Galaxy deepTools (Ramírez et al., 2014).

Transcriptomic analysis

De novo transcriptomes were assembled using StringTie 1.3.3 software (Pertea et al., 2015). 12 RNA-seq datasets were used for *de novo* transcriptome assembly (three replicates of each cell type: female B cells, male B cells, female T cells and male T cells). Using Bedtools (Quinlan and Hall, 2010), the coordinates of the identified transcripts were overlapped with VM-IAP coordinates to identify transcripts initiated or terminated within VM-IAPs. Only VM-IAPs that overlapped transcripts in at least three biological replicates representing the same cell type and sex were further analyzed.

Inheritance analysis

The effect of parental methylation level on offspring methylation level was analyzed using REML-fitted linear mixed-effects models (LMMs) in R via the lmer() function in the lme4 package (Bates et al., 2015). The bisulfite pyrosequencing methylation levels of all individuals used for this analysis were averaged across the first four CpGs of the IAP LTR and run through a logit transformation before feeding into the model. The LMMs for each VM-IAP included maternal methylation level, paternal methylation level, and sex as fixed effects. Breeding pair as well as litter nested within breeding pair were treated as random intercept effects, accounting for the non-independence of siblings and littermates, respectively. Interaction between maternal and paternal methylation levels was originally assessed but no significant interaction was found. Parameter estimates, standard errors, and t values are reported in Figure S5B. By default, the reference intercept is selected alphabetically, in this case representing the estimate for female methylation level. To evaluate the significance of the fixed effects, p values were generated using the Satterthwaite approximation for degrees of freedom, applied by the lmerTest package in R (Kuznetsova et al., 2017). To account for multiple testing, q-values were generated using the Benjamini-Hochberg correction with a false discovery rate of 0.05 (Benjamini and Hochberg, 1995). To assess effect sizes, semi-partial R^2 values for each fixed effect were calculated using the r2beta() function from the r2glmm package in R (Jaeger et al., 2017). For the repeated inheritance experiment on VM-IAP$_{Gm13849}$, the mean methylation level of littermates was calculated for the first litter of five highly methylated and five lowly methylated mothers. Significance was assessed with a one-sided unpaired t test using GraphPad Prism.

Assembly of ERV coordinates

The RepeatMasker database was downloaded from the UCSC Table Browser to determine the genomic coordinates of ERV fragments (Karolchik et al., 2004). ERV fragments were separated into four groups according to RepeatMasker annotation (ERV1, ERVL, ERVK and IAPs) and assembled using Bedtools (Quinlan and Hall, 2010). The structure of ERV insertions was determined based on RepeatMasker insertion fragment annotation.

DATA AND SOFTWARE AVAILABILITY

The accession number for the WGBS, WGoxBS, and RNA-seq datasets used in this study is GEO: GSE94676. Cell purification and DNA/RNA extraction protocols as well as data processing pipelines are available at the above accession site. The R code used for the algorithms and computation analyses in this study has been collated into the following Github file: https://github.com/AFS-lab/Kazachenka-Bertozzi-et-al-2018.

Cell

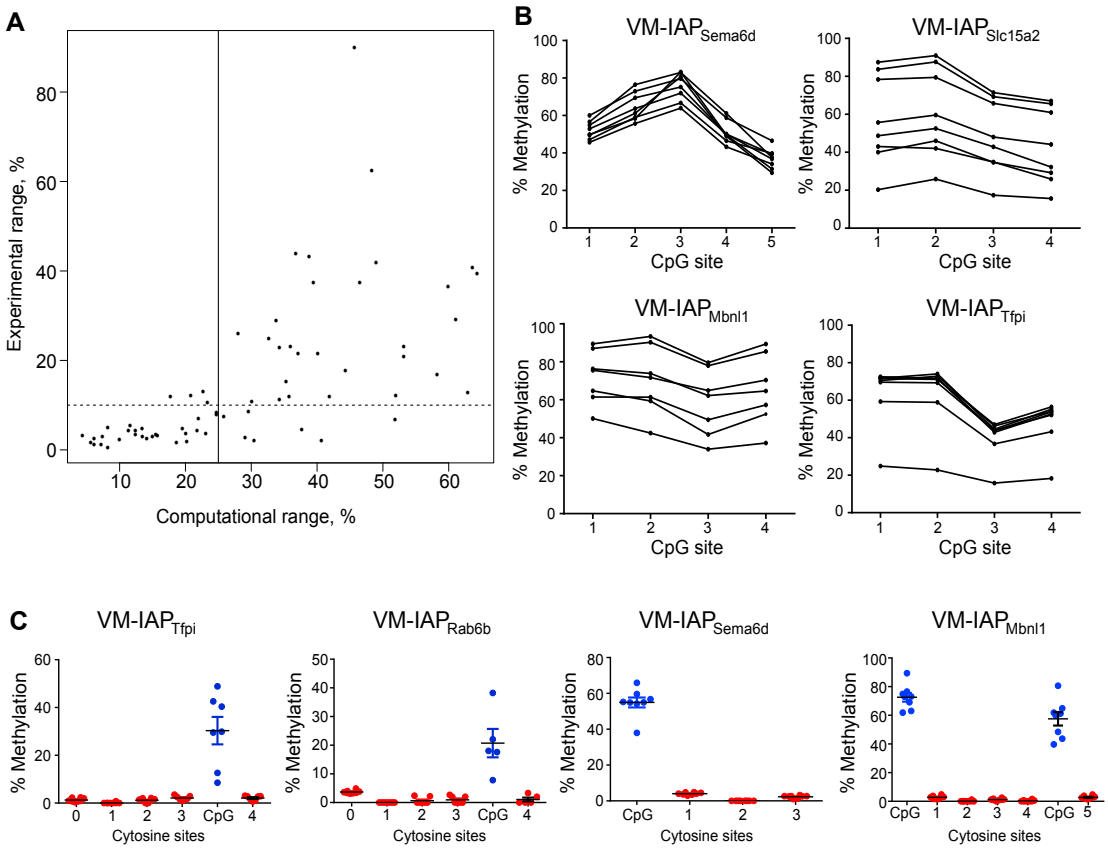


Figure S1. Validation and Characterization of VM-IAPs, Related to Figure 1

(A) Validation of methylation variation threshold used for the genome-wide screen. Each dot represents an IAP. Experimental range represents the difference between average methylation levels of the most highly and lowly methylated individuals, identified via bisulfite pyrosequencing. The computational and experimental ranges are correlated (two-tailed Pearson). The vertical line defines the threshold used for the genome-wide screen and the dotted horizontal line represents the upper range of experimental error associated with pyrosequencing.

(B) Inter-individual methylation variation at VM-IAP$_{Tfpi}$, VM-IAP$_{Mbnl1}$, VM-IAP$_{Slc15a2}$, and VM-IAP$_{Sema6d}$, tested in pure B cell populations extracted from eight individuals, confirming the relationship between the BLUEPRINT experimental data generated from whole C57BL/6J tissues.

(C) Non-CG methylation at four VM-IAPs, assessed by bisulfite pyrosequencing. Each dot represents an individual. Non-CG cytosines are shown as red dots and CpGs are shown as blue dots.

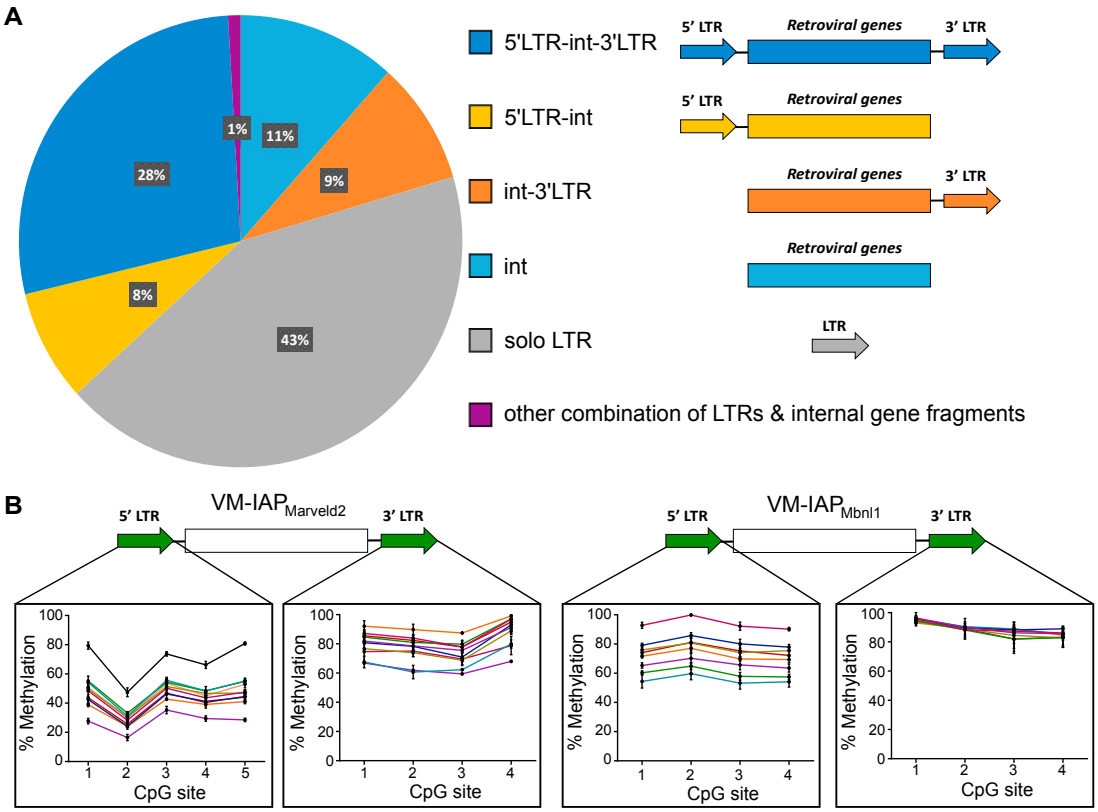
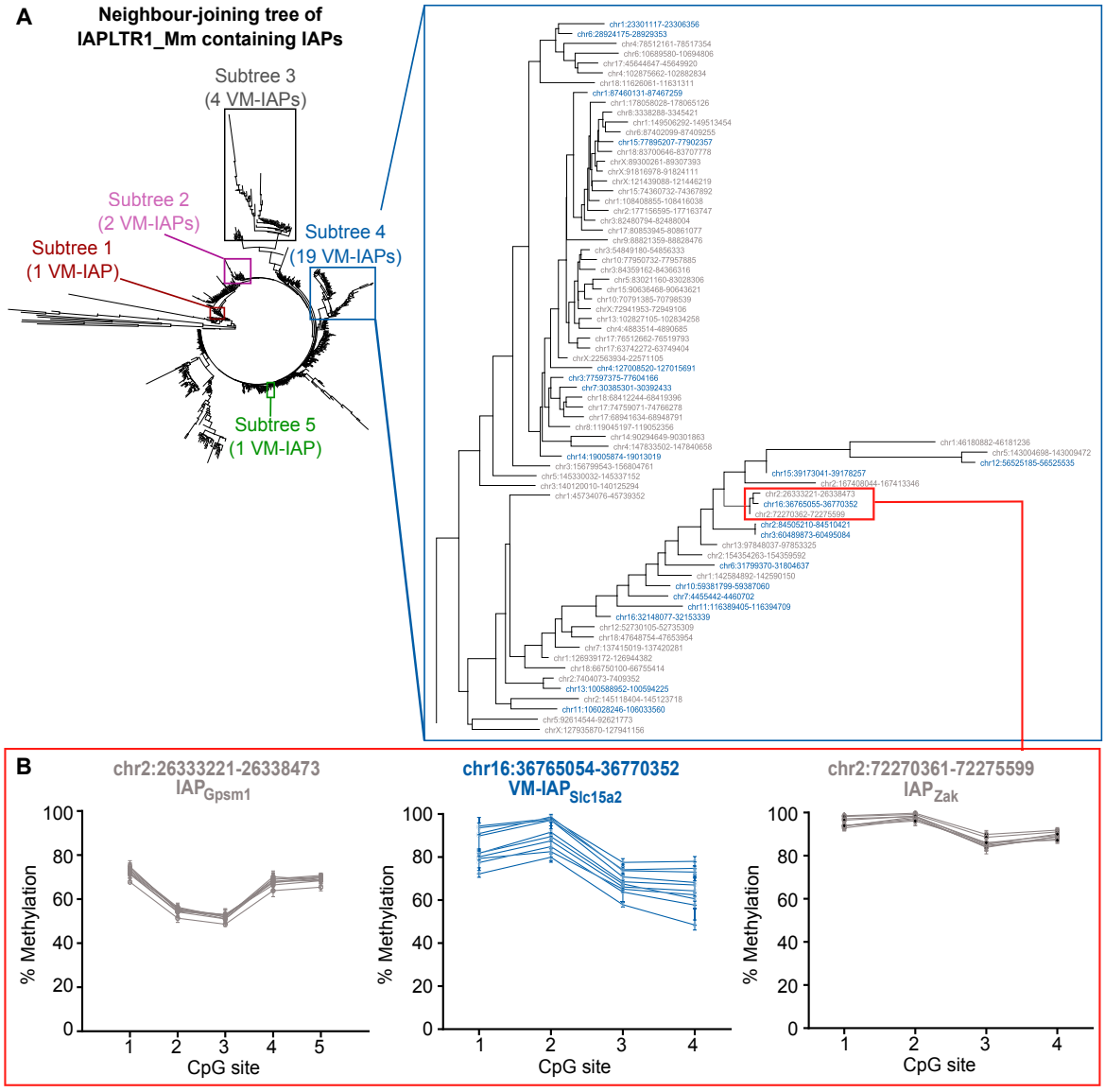


Figure S2. Structural Analysis of VM-IAPs, Related to Figure 1

(A) Pie chart distribution of all IAPs in the C57BL/6J genome based on IAP structure, where "int" refers to internal retroviral sequence.

(B) Examples illustrating that LTRs behave independently of each other. Bisulphite pyrosequencing of 5′ and 3′ LTRs of VM-IAP$_{Marveld2}$ and VM-IAP$_{Mbnl1}$. The same individuals were used to assess both LTRs, color-coded accordingly.



C Co-variation analysis

	VM-IAP$_{Bmf}$	VM-IAP$_{Tfpi}$	VM-IAP$_{Mbnl1}$	VM-IAP$_{Rnf157}$	VM-IAP$_{Marveld2}$	VM-IAP$_{Gm13849}$
VM-IAP$_{Bmf}$		0.005[1] 0.980[2]	0.092 0.609	0.211 0.238	0.148 0.413	-0.139 0.439
VM-IAP$_{Tfpi}$	0.005 0.980		-0.087 0.630	0.360 0.040	0.061 0.737	-0.263 0.140
VM-IAP$_{Mbnl1}$	0.092 0.609	-0.087 0.630		0.316 0.073	0.233 0.192	0.214 0.231
VM-IAP$_{Rnf157}$	0.211 0.238	0.360 0.040	0.316 0.073		0.071 0.693	0.048 0.792
VM-IAP$_{Marveld2}$	0.148 0.413	0.061 0.737	0.233 0.192	0.071 0.693		0.215 0.229
VM-IAP$_{Gm13849}$	-0.139 0.439	-0.263 0.140	0.214 0.231	0.048 0.792	0.215 0.229	

[1]Pearson correlation coefficient
[2]P-value (Bonferroni-adjusted α value = 0.008, n = 33 mice)

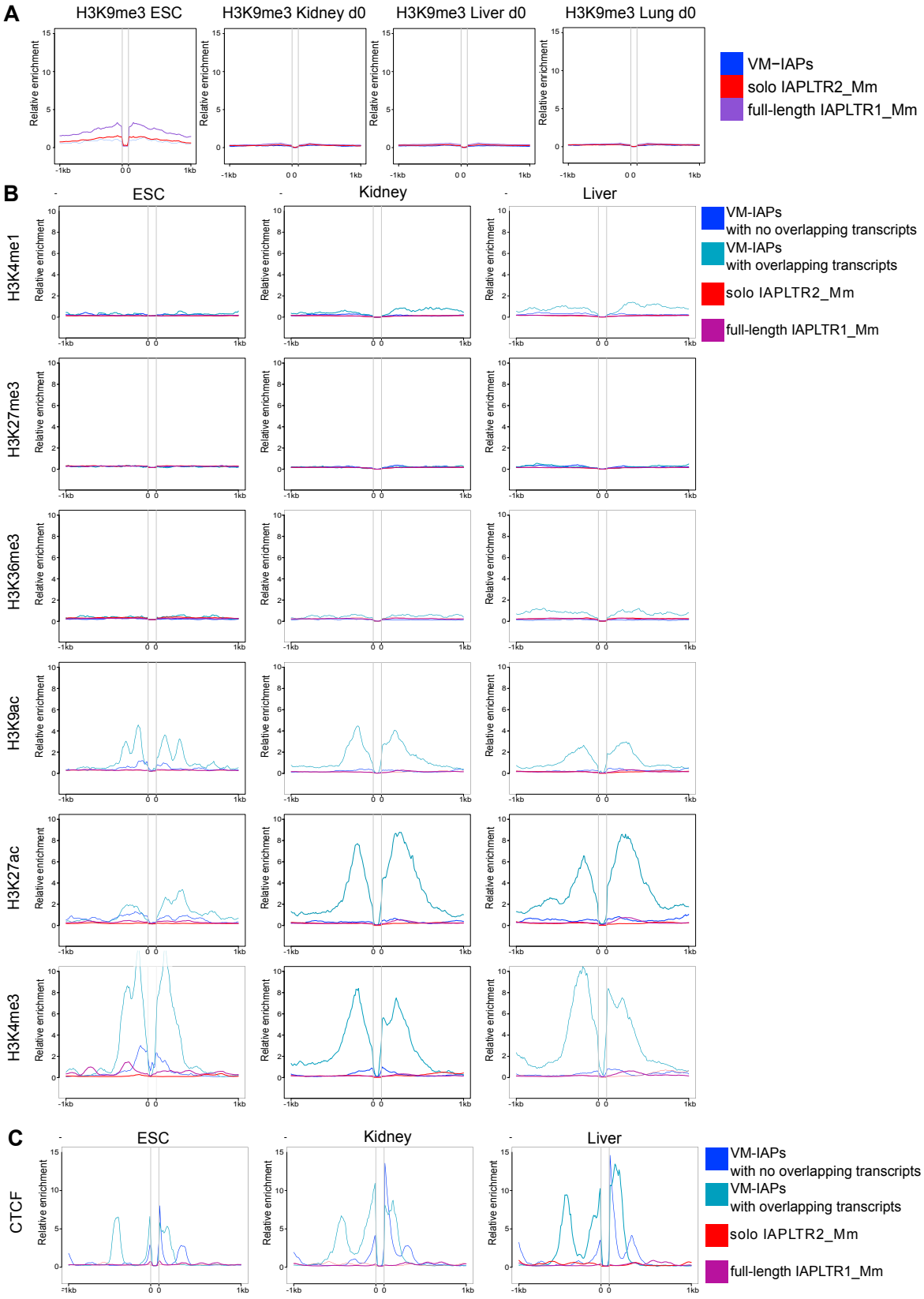
(legend on next page)

Figure S3. Genetic Sequence and Co-variation Analyses of VM-IAPs, Related to Figure 2

(A) Neighbor-joining tree for IAPLTR1_Mm elements, made using Geneious software. VM-IAPs are distributed across 5 subtrees, shown in boxes. Subtree 4, containing the most VM-IAPs, is shown in more detail, with VM-IAP coordinates highlighted in blue.

(B) IAP_{Zak} and IAP_{Gpsm1}, closely related to $VM\text{-}IAP_{Slc15a2}$ by sequence, do not exhibit inter-individual methylation variation upon bisulphite pyrosequencing (n = 10).

(C) Normalized correlation matrix of methylation levels across 6 VM-IAPs in 33 mice.



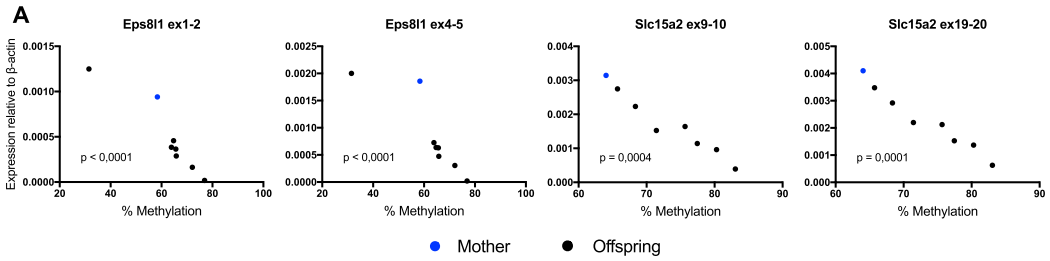
(legend on next page)

Figure S4. Epigenetic Profiles of VM-IAP Flanking Regions, Related to Figure 3 and 4

(A) Relative H3K9me3 enrichment profiles of VM-IAP flanking genomic regions in ESCs, kidney, liver and lung. VM-IAPs are not flanked by H3K9me3-enriched regions and are equivalent to full-length IAPLTR1_Mm and solo IAPLTR2_Mm IAPs. All IAPs were fit to 100bp, shown as the space between the two 0's. 0's represent the start and end coordinates of IAPs. ChIP-seq datasets were downloaded from ENCODE.

(B) Epigenetic profiles of regions flanking VM-IAPs, separated by presence or absence of overlapping *de novo* assembled transcripts. Highly methylated full length IAPs of the IAPLTR1_Mm subclass and solo LTRs of the IAPLTR2_Mm subclass serve as controls. All IAPs were fit to 100bp, shown as the space between the two 0's. 0's represent the start and end coordinates of the IAPs.

(C) Relative CTCF enrichment profiles of regions flanking VM-IAPs, separated by presence or absence of overlapping *de novo* assembled transcripts.



B

Linear mixed-effects models: fixed effect output

Predictor	Estimate	Standard error	t value	R^{2*}	p-value	q-value	n
VM-IAP$_{Marveld2}$							183
Maternal % methylation†	-0.016	0.085	-0.190	0	0.849	0.849	
Paternal % methylation†	0.023	0.081	0.290	0.001	0.776	0.847	
Sex	0.204	0.088	2.320	0.026	0.022	0.113	
VM-IAP$_{Bmf}$							278
Maternal % methylation†	0.119	0.096	1.240	0.021	0.240	0.540	
Paternal % methylation†	-0.038	0.152	-0.250	0.001	0.800	0.847	
Sex	0.071	0.070	1.010	0.003	0.310	0.569	
VM-IAP$_{Rnf157}$							213
Maternal % methylation†	0.032	0.093	0.350	0.001	0.730	0.847	
Paternal % methylation†	0.068	0.073	0.940	0.008	0.360	0.569	
Sex	0.367	0.082	4.490	0.078	1.20e-5	**2.16e-4**	
VM-IAP$_{Tfpi}$							264
Maternal % methylation†	0.094	0.062	1.520	0.020	0.171	0.513	
Paternal % methylation†	0.025	0.077	0.320	0.001	0.757	0.847	
Sex	0.152	0.121	1.260	0.006	0.210	0.540	
VM-IAP$_{Mbnl1}$							265
Maternal % methylation†	0.117	0.114	1.030	0.012	0.323	0.569	
Paternal % methylation†	0.241	0.099	2.430	0.068	0.048	0.173	
Sex	-0.039	0.065	-0.600	0.001	0.547	0.757	
VM-IAP$_{Gm13849}$							262
Maternal % methylation†	0.301	0.080	3.780	0.166	0.004	**0.036**	
Paternal % methylation†	-0.091	0.098	-0.920	0.011	0.379	0.569	
Sex	0.142	0.063	2.260	0.017	0.025	0.113	

*R^2: semi-partial R^2 for each fixed effect.
†Logit-transformed.

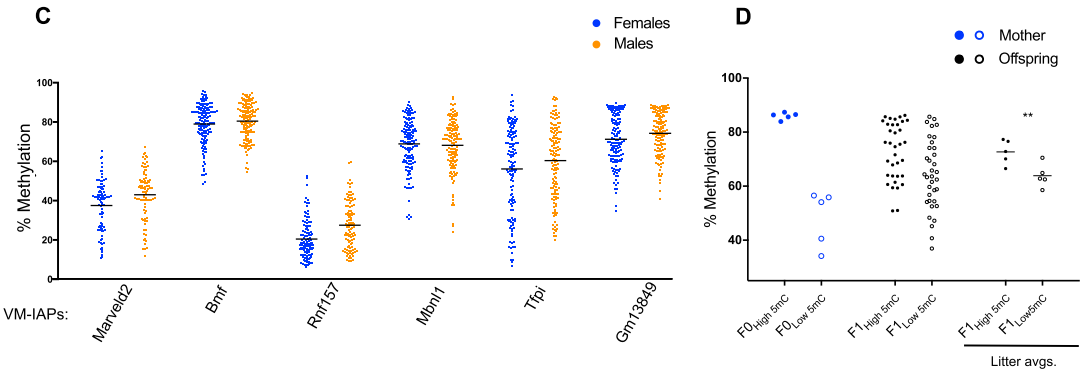


Figure S5. Transcriptional and Statistical Results of VM-IAP Inheritance Studies, Related to Figures 5 and 6

(A) The inverse correlation between expression of *Eps8l1* (exons 1-2 and 4-5) and VM-IAP$_{Eps8l1}$ methylation is recapitulated in the F1 generation (two-tailed Pearson). This is also observed for expression of *Slc15a2* (exons 9-10 and 19-20) and VM-IAP$_{Slc15a2}$ methylation. Expression was quantified in spleen by qPCR and shown relative to housekeeping gene *β-actin*. Each dot represents a different individual and maternal expression is shown in blue.

(B) Linear mixed-effects models (LMMs) of offspring methylation for six VM-IAPs using the lmerTest package in R. Maternal methylation level, paternal methylation level, and sex were treated as fixed effects. Breeding pair and litter were treated as random effects. Output for the fixed effects is presented. Raw p values were

(legend continued on next page)

generated using the Satterthwaite approximation for degrees of freedom. Adjusted p values (q-values) were generated using the Benjamini & Hochberg correction to account for multiple testing.

(C) Methylation levels of offspring used to fit LMMs, separated by sex. A significant sex effect is observed for VM-IAP$_{Rnf157}$ (p value: 1.20e-5; q-value: 2.16e-4). Average methylation levels are shown in black bars.

(D) Validation of the observed maternal effect on offspring methylation levels at VM-IAP$_{Gm13849}$. The average methylation levels of the first litters born to five highly methylated and five lowly methylated C57BL/6J females were assessed from ear notches via bisulfite pyrosequencing (one-sided t test; p value: 0.0069; n = 5 litters per group).

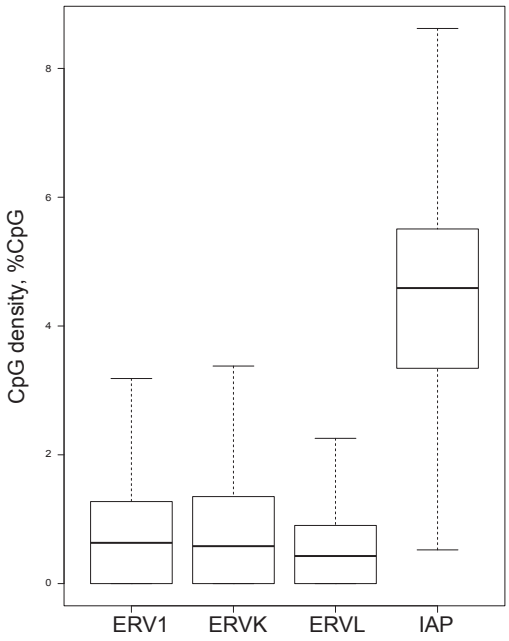


Figure S6. Classification of ERV Subtypes by CpG Density, Related to STAR Methods

ERV LTR CpG density, separated by subtype. DNA sequences for all ERV LTR regions were extracted and sorted according to RepeatMasker annotation. CpG density was calculated as the percentage of CpGs to base pair length of the LTR region.

Seminars in Cell and Developmental Biology 97 (2020) 93–105

Contents lists available at ScienceDirect

Seminars in Cell & Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Metastable epialleles and their contribution to epigenetic inheritance in mammals

Tessa M. Bertozzi, Anne C. Ferguson-Smith*

Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK

ARTICLE INFO

Keywords:
Metastable epialleles
Agouti viable yellow
DNA methylation
Retrotransposon
Epigenetic inheritance

ABSTRACT

Many epigenetic differences between individuals are driven by genetic variation. Mammalian metastable epialleles are unusual in that they show variable DNA methylation states between genetically identical individuals. The occurrence of such states across generations has resulted in their consideration by many as strong evidence for epigenetic inheritance in mammals, with the classic A^{vy} and $Axin^{Fu}$ mouse models – each products of repeat element insertions – being the most widely accepted examples. Equally, there has been interest in exploring their use as epigenetic biosensors given their susceptibility to environmental compromise. Here we review the classic murine metastable epialleles as well as more recently identified candidates, with the aim of providing a more holistic understanding of their biology. We consider the extent to which epigenetic inheritance occurs at metastable epialleles and explore the limited mechanistic insights into the establishment of their variable epigenetic states. We discuss their environmental modulation and their potential relevance in genome regulation. In light of recent whole-genome screens for novel metastable epialleles, we point out the need to reassess their biological relevance in multi-generational studies and we highlight their value as a model to study repeat element silencing as well as the mechanisms and consequences of mammalian epigenetic stochasticity.

1. Introduction

Inter-individual phenotypic and epigenetic variation is most often explained by underlying genetic polymorphism. There is evidence, however, that in certain instances genetically identical individuals can differ epigenetically. The mechanisms driving such differences are poorly understood and likely involve both external (environmental) factors as well as intrinsic stochastic processes. There is considerable interest in determining the extent to which epigenetic information can be passed on from one generation to the next, as this challenges the dogma dictating that heritable traits are strictly conferred by the sequence of DNA transmitted from parent to offspring. Epigenetic inheritance across generations has convincingly been shown to occur in a number of non-mammalian model organisms. This has been reviewed elsewhere [1–3]. In contrast, this type of inheritance is rare in mammals due to the extensive genome-wide epigenetic reprogramming that takes place during mammalian development. Where it does occur, the driving mechanisms remain poorly understood. Metastable epialleles are frequently cited as the best example of this phenomenon in mammals. This review, focused solely on mammalian biology, will explore and interpret the literature to date regarding these unusual loci.

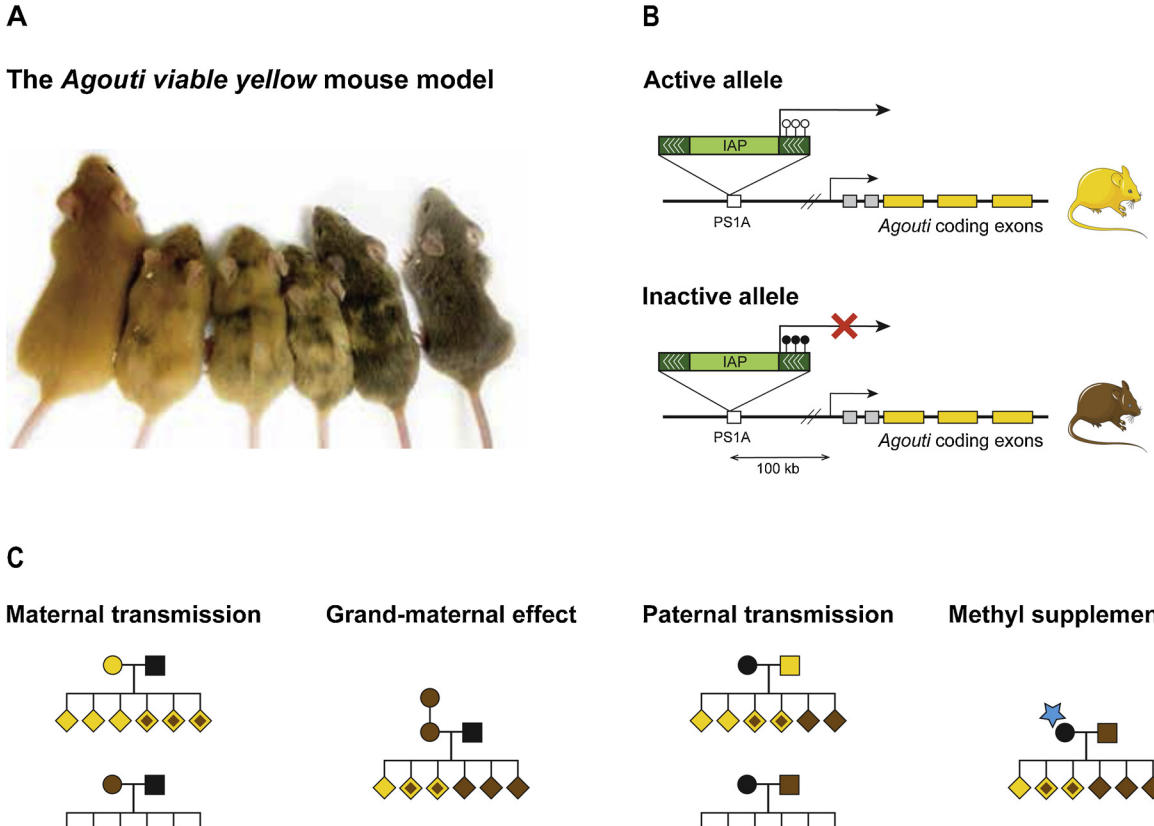
2. Classic metastable epialleles: A^{vy} and $Axin^{Fu}$

In order for the definition of metastable epialleles to become clear, let us first consider two classic examples in the mouse: the *Agouti viable yellow* (A^{vy}) and *Axin fused* ($Axin^{Fu}$) alleles. Both exhibit ectopic gene expression due to intracisternal A-particle (IAP) insertions [4,5]. IAPs are repetitive elements of the Class II endogenous retrovirus (ERV) family, their structure characterized by protein-coding sequences flanked by 5′ and 3′ long terminal repeats (LTRs) [6]. LTRs contain regulatory sequences that can act as host gene promoters and enhancers, making retrotransposition events potential drivers of evolutionary change (reviewed in [7]).

The *Agouti viable yellow* (A^{vy}) allele arose from the spontaneous insertion of an IAP into pseudoexon1A (PS1A) of the *Agouti* coat colour gene locus [4,8,9] (Fig. 1A and B). PS1A is located approximately 100 kb upstream of the *Agouti* coding exons (Fig. 1B). Wild type *Agouti* is normally expressed transiently from a hair cycle-specific promoter, producing a paracrine signalling peptide that yields a yellow band on a black hair [10]. This tightly controlled expression pattern is responsible for brown wild-type ‘agouti’ fur. The Agouti peptide can also interfere with metabolic pathways and has been linked to obesity, glucose intolerance, and tumourigenesis [11,12]. In A^{vy} mice, a cryptic promoter

* Corresponding author.
E-mail address: afsmith@gen.cam.ac.uk (A.C. Ferguson-Smith).

https://doi.org/10.1016/j.semcdb.2019.08.002
Received 22 November 2018; Received in revised form 15 August 2019; Accepted 20 August 2019
Available online 21 September 2019



(caption on next page)

in the IAP LTR drives constitutive ectopic *Agouti* expression [4] (Fig. 1B). This produces a mouse with a completely yellow coat as well as adult-onset obesity and diabetes (reviewed in [11]). The A^{vy} allele is part of a series of dominant *Agouti* alleles brought about by IAP insertions, suggesting the *Agouti* gene locus may be particularly prone to insertional mutagenesis. Some of these alleles, such as A^{iapy} and A^{hvy}, cause similar phenotypes to A^{vy} [13,14]. These additional dominant *Agouti* alleles have not been studied as extensively as A^{vy} and will not be discussed further.

The *Axin fused* ($Axin^{Fu}$) allele resulted from an IAP insertion in the sixth intron of the *Axin* gene [5,15]. The Axin protein is involved in the regulation of embryonic axis formation by inhibiting Wnt signaling [16]. The intragenic position of the IAP causes aberrant transcription of *Axin* downstream exons, producing an atypical protein. The resulting truncated Axin interferes with axial patterning and results in the development of a distinctive kinked tail [5,17]. Of note, while the $Axin^{Fu}$-associated IAP is intragenic, the one associated with the A^{vy} locus is 100 kb upstream of the affected exons, indicating that retrotransposon-

Fig. 1. The *Agouti viable yellow* (A^{vy}) locus.
(A) Genetically identical A^{vy} mice display a range of coat colours from yellow to pseudoagouti, including varying levels of mottling in between. The mice shown are on a C57BL/6J background and are not age-matched siblings.
(B) The A^{vy} allele is characterized by the presence of a contra-oriented IAP insertion (directionality shown with white arrows) in pseudoexon1A (PS1A) 100 kb upstream of the *Agouti* coding exons. The IAP is variably methylated between individuals and drives constitutive *Agouti* expression when unmethylated, leading to yellow fur. Full methylation results in pseudoagouti fur and partial methylation gives rise to mottling (see panel A).
(C) Maternal coat colour phenotype influences the coat colour distribution in A^{vy}/a offspring in the C57BL/6 genetic background. A grand-maternal effect is observed when the A^{vy} allele is transmitted through two generations of pseudoagouti females, resulting in a more severe phenotypic shift. No inheritance is observed upon paternal transmission on this genetic background (adapted from [18]). *In utero* dietary methyl supplementation shifts A^{vy} coat colour towards pseudoagouti [9]. In these experiments, mice carrying the A^{vy} allele were bred to congenic a/a mice homozygous for the *Agouti* null allele (a). Pedigrees: circle—female; square—male; diamond—sex unspecified; a/a offspring are not included.
(D) DNA methylation of the A^{vy} allele in gametes and blastocysts upon maternal and paternal transmission suggests reprogramming after fertilization. Methylation levels in sperm and oocytes reflect the coat colour phenotype and somatic methylation levels of the individual, while blastocysts are largely unmethylated regardless of parental phenotype. This is consistent with erasure of DNA methylation during preimplantation stages. The diagrams are constructed based on clonal bisulphite sequencing data from [52]. The methylation state of blastocysts produced by yellow A^{vy}/a dams has not been studied (depicted as a question mark).

mediated gene regulation can occur both locally and from a distance.

Unlike most mutations, A^{vy} and $Axin^{Fu}$ mice show variable penetrance and expressivity despite genetic homogeneity. The coat colour of individual inbred A^{vy} mice ranges from completely yellow to seemingly wild-type agouti (termed *pseudoagouti*), including varying intermediate degrees of mottled patterns [18] (Fig. 1A). Likewise, tail morphologies in inbred $Axin^{Fu}$ mice span from straight to severely kinked [17]. In both cases, the mechanism underlying the continuous phenotypic spectra is epigenetic in origin, with DNA methylation at the IAPs inversely correlated with expressivity. Hypomethylation at the IAP LTR thus corresponds to increased allelic expression, and vice versa [17,18]. This is highly unusual for IAPs, the vast majority of which are heavily methylated [19]. Remarkably, the full span of phenotypes can be observed within a single litter regardless of parental phenotype, illustrating the instability of their epigenetic state after passage through the germline [8,15].

The term *metastable epiallele* should make more sense now. The word *metastable* was first used in this context by plant biologists to describe alleles whose epigenetic state is capable of switching between active and repressed states from one generation to another [20–22]. Adapted for mammals by Emma Whitelaw and colleagues in 2002, the term *metastable epiallele* is intended to highlight (1) the epigenetic basis for the phenotypes associated with these alleles and (2) the apparent stochasticity of their epigenetic state [23]. In practice, this has translated to methylation variation between genetically identical individuals and, importantly, consistency in methylation levels within a single individual. While the methylation level of the IAPs associated with A^{vy} and $Axin^{Fu}$ varies across different mice, it is constant across different tissues of a single mouse, suggesting the methylation state is established early in development before tissue differentiation and maintained mitotically thereafter [9,24]. This differentiates metastable epialleles from *differentially methylated regions* (DMRs), a broader term used to designate differential but invariant methylation between biological samples, which could be cells, tissues, or individuals depending on the context.

Methylation consistency across tissues at the A^{vy} locus may seem at odds with the variegated patches of yellow and agouti fur observed on mottled mice. One might have expected pelts of intermediately methylated individuals to display a consistent intermediate pigmentation. DNA methylation, however, is dichotomous. A single CpG site is either methylated or unmethylated. Hence, the evident inter-individual range of methylation results from different proportions of methylated alleles in a cell population. Following from that, the proportion of methylated cells at A^{vy} is likely determined in a probabilistic fashion before germ layer differentiation and propagated mitotically throughout the body as it develops [9,11]. This would result in an approximately equal methylation percentage across tissues but would allow for local patches of cells to be different depending on the methylation state of their clonal origin. This is reminiscent of the black and orange mosaic fur pigmentation observed in tortoiseshell cats due to X-chromosome inactivation [25]. Given that cellular development and proliferation differ between cell types, it is perhaps more accurate to think of metastable epialleles as loci that display a substantial correlation in methylation between tissues rather than identical intra-tissue methylation.

3. Assessing the prevalence of metastable epialleles genome-wide

3.1. Murine metastable epialleles

As discussed above, retroelement insertions play a key role in the unique behaviour of the A^{vy} and $Axin^{Fu}$ loci. Considering almost half of the mouse genome is made up of repetitive elements, it is perhaps unsurprising that other metastable epialleles have been identified. While A^{vy} and $Axin^{Fu}$ were discovered decades ago due to their striking visual phenotypes [8,15], the identification of additional candidates has relied on using the genetic and epigenetic features of these classic loci to develop genome-wide screens and search algorithms. The third metastable epiallele to be identified was discovered by inspecting C57BL/6J cDNA databases in the hopes of finding transcripts containing IAP LTR sequences [26]. One such sequence contained a contra-oriented IAP element in the sixth intron of the *Cdk5rap1* gene. Much like the A^{vy} and $Axin^{Fu}$ loci, IAP LTR methylation was inversely correlated with expression of the aberrant transcript initiating from the 5′ LTR. The inter-individual methylation range of this new candidate, named $Cabp^{IAP}$, is much narrower than those observed at A^{vy} and $Axin^{Fu}$, and no identifiable phenotype is associated with its epigenetic variability [26].

The advent of high-throughput sequencing in the past decade has enabled larger-scale screens for metastable epialleles. The first attempt searched the mouse genome using genome-wide expression microarray data. Transcripts exhibiting wide-ranging inter-individual variation and low-ranging inter-tissue variation were selected as candidates in an attempt to capture the expression pattern observed for the *Agouti* gene in A^{vy} mice [27]. Only two loci (*Dnajb1* and *Glcci1*) were analysed in depth and though they showed inter-individual methylation differences, neither exhibited methylation-associated expression.

The second attempt screened for IAP insertions with promoter activity. The study identified retrotransposons near mRNA promoters associated with H3K4me3, an activating histone modification [28]. This enriched for active IAP LTR promoters and resulted in a set of 143 candidate regions, from which 13 were selected for experimental validation. Only three of these were found to exhibit significant methylation variation between individuals [28]. Also focusing on repeat elements, Faulk and colleagues recognized that the IAPs associated with A^{vy} and $Cabp^{IAP}$ both belong to the IAPLTR1_Mm subclass. They showed that IAPLTR1_Mm elements cluster into three clades, with the largest one containing the most conserved elements. A^{vy} and $Cabp^{IAP}$ segregated together in a separate smaller clade. Based on a limited selection of seven loci per clade, they provide preliminary evidence that the younger clades are more lowly methylated and display greater inter-individual methylation ranges [29]. These clades are likely enriched for

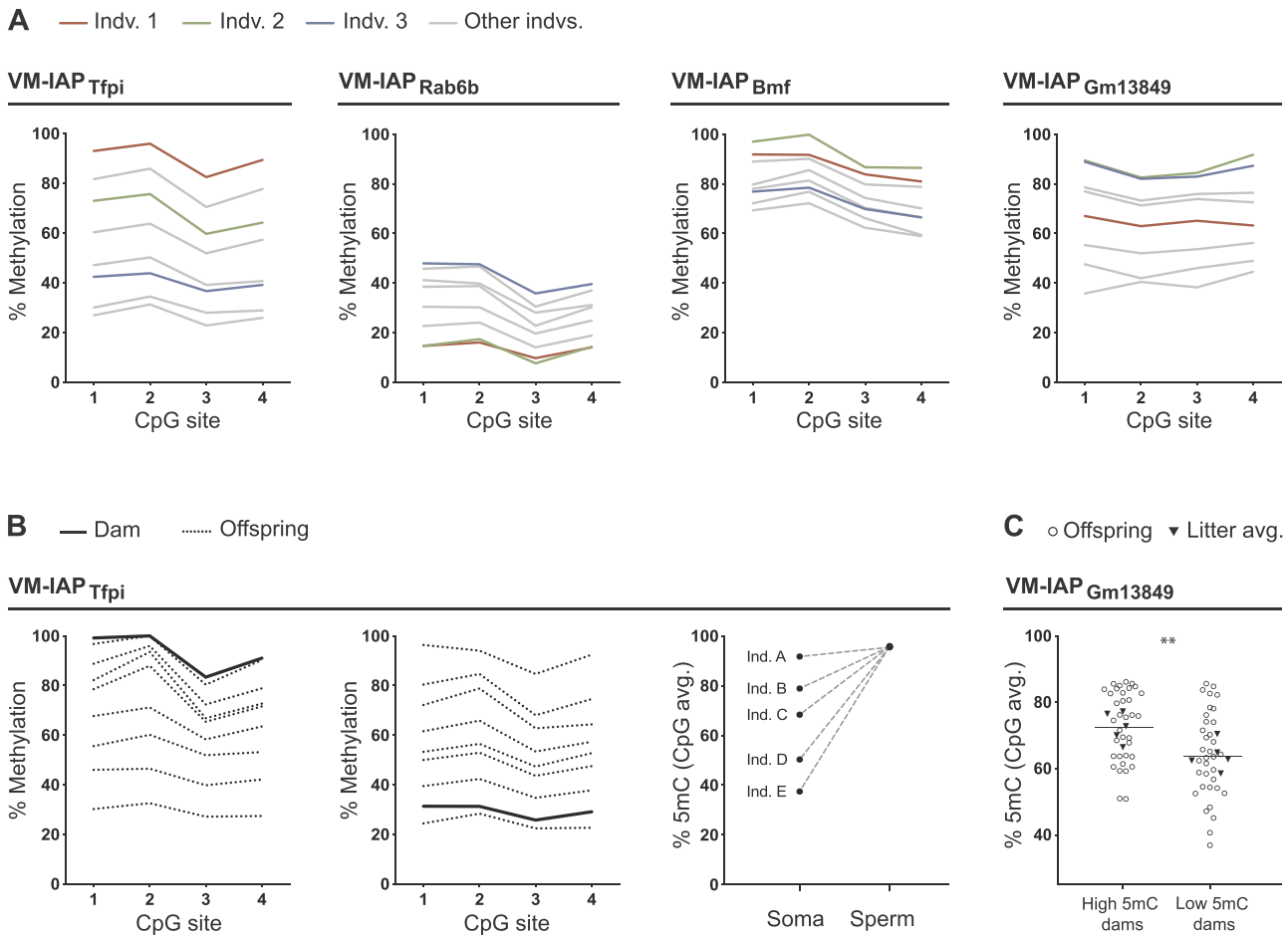


Fig. 2. Epigenetic variability and inheritance at VM-IAPs.
(A) VM-IAPs exhibit inter-individual methylation variation (four examples are shown). The methylation level at VM-IAPs is locus-specific, whereby separate loci show different methylation levels within a single mouse.
(B) The full range of VM-IAP methylation variation is predictably reconstructed from one generation to the next regardless of maternal methylation level. In contrast to the A^{vy} and $Axin^{Fu}$ loci, VM-IAPs are hypermethylated in mature sperm irrespective of somatic methylation levels (e.g. VM-IAP$_{Tfpi}$). Adapted from [31].
(C) Maternal epigenetic inheritance at VM-IAP$_{Gm13849}$. On average, offspring born to highly methylated dams show higher methylation levels than offspring born to lowly methylated dams. These results are from a replicate cohort. Statistics: one-sided *t*-test on litter averages, n = 5 dams per group. Adapted from [31].

metastable epialleles.

Taking an unbiased approach, Oey and colleagues performed whole genome bisulphite sequencing on five A^{vy} mice and identified 356 regions showing inter-individual methylation variation, including 55 ERV-overlapping regions [30]. Four of these were experimentally validated. Perhaps the most valuable aspect of this study was the whole genome sequencing analysis of two A^{vy} mice with different coat colours. Only 32 single nucleotide variants were detected in coding sequences and no mutations were found near the A^{vy} allele, contesting arguments suggesting that the phenotypic variation observed among A^{vy} littermates is due to genetic variation. However, as explained in Section 4.2, genetic background has an effect on the distribution of A^{vy} phenotypes in a parent-of-origin–specific manner.

The most recent screen for metastable epialleles was conducted by our team. We carried out a genome-wide screen identifying variably methylated IAPs (VM-IAPs) in the C57BL/6J genome [31]. Extensive characterization of the resulting candidates revealed both similarities and differences to A^{vy} and $Axin^{Fu}$. Much like the classic alleles, VM-IAP methylation levels are variable between individuals but consistent across tissues within a single individual (Fig. 2A). In addition, inter-individual methylation variation at VM-IAPs is reconstructed from generation to generation regardless of parental methylation level (Fig. 2B). However, the transcriptional effects and inheritance patterns observed at VM-IAPs, discussed in more depth in Sections 4 and 7, indicate that properties associated with A^{vy} and $Axin^{Fu}$ cannot necessarily be extrapolated to other metastable epialleles. To avoid confusion, the term VM-IAP will be used in this review when discussing findings specific to the regions identified and characterized in this screen.

Although the loci discussed here are endogenous and naturally occurring, a chimeric long interspersed nuclear element (LINE) retrotransposon of the L1 subclass was recently experimentally inserted into the *Axin* gene, inducing the kinked tail phenotype with variable penetrance much like the spontaneous $Axin^{Fu}$ insertion [32]. Unlike $Axin^{Fu}$, the $Axin^{cL1}$ mutation is not associated with variable methylation levels. Additional experimental manipulations of this sort will be of great comparative value in determining the mechanisms underlying metastability.

3.2. Metastable epialleles in humans

The identification of human metastable epialleles is a very challenging task due to the extensive genetic variation present in human populations. It is nevertheless an important endeavour in order to assess the extent to which they contribute to human phenotypic variation, to understand the roles of the human repeat genome, and to evaluate the relevance of using murine metastable epialleles as models to study epigenetic variation in humans. While analysing monozygotic (MZ)

twin cohorts can aid in overcoming some of the challenges associated with genetic variation, a recent cautionary study describes the phenomenon of "epigenetic supersimilarity" between human MZ twins and highlights their non-equivalence to isogenic mice that do not originate from the same zygote [33].

An alternative strategy to control for genetic differences in human populations has been to use a large number of genetically diverse methylomes. This approach has given rise to a growing list of putative human metastable epialleles, some of which are sensitive to environmental factors such as maternal nutrition and season of conception [34–37]. The most recent of these studies argues that the definition of metastable epialleles should be relaxed to include variably methylated regions that are susceptible to genetic influence, at least in the human context where this issue is unavoidable [37]. Under this framework, additional sequence-dependent candidates identified in the past year can be added to the list of potential human metastable epialleles [38,39].

Interestingly, although IAP elements do not exist in humans, the bordering regions of putative human metastable epialleles appear to be enriched for transposable elements of the ERV and LINE families [36,37]. While some of the murine screens were specifically restricted to transposable elements based on the presence of IAPs at the A^{vy} and $Axin^{Fu}$ loci [28,29,31], others were unbiased and still found enrichment for repeats [27,30]. Taken together, the mouse and human studies indicate that repeat elements play an important and conserved role in the establishment of inter-individual epigenetic variability. It is possible that metastable epialleles are a product of conflicting interactions between activating factors recruited to insertion sites and repeat repressive modifiers, an idea we will return to in Section 5. This does not preclude, however, the possibility of metastable epialleles at unique non-repetitive regions maintained by distinct mechanisms. The human studies suggest these exist, but a systematic screen for metastability at unique regions has not, to our knowledge, been conducted in the mouse.

4. Metastable epialleles as models of epigenetic inheritance

4.1. Partial inheritance of parental epigenetic state

Epigenetic inheritance across generations is one of the most striking properties of A^{vy} and $Axin^{Fu}$ mice. In the case of A^{vy}, maternal (but not paternal) coat colour phenotype affects the range of phenotypes observed in the offspring; the coat colour distribution of offspring born to yellow mothers is shifted towards yellow compared to that of offspring born to pseudoagouti mothers [40,41] (Fig. 1C). These experiments were conducted using inbred mouse strains, effectively eliminating the possibility of genetically mediated effects [18,40,41]. In light of the increased incidence of obesity observed in yellow mice, maternal inheritance of coat colour at A^{vy} was originally attributed to metabolic differences in the intrauterine environments of developing embryos [40]. Elegant embryo transfer experiments showed that this is not the case. Transferring fertilized oocytes from yellow dams to black foster mothers not carrying the A^{vy} allele produces offspring with the same coat colour distribution as offspring born to yellow dams without embryonic intervention [18]. This confirms that the transmission of maternal coat colour to the next generation is an epigenetic process rather than an environmental one. The same study reported a grand-maternal effect at A^{vy}: transmission of the allele through two generations of pseudoagouti dams appeared to cause a greater shift towards pseudoagouti than transmission through a single generation (Fig. 1C). It is unknown whether passage through a third or fourth generation of pseudoagouti females produces a further cumulative effect.

In contrast to A^{vy}, the $Axin^{Fu}$ allele exhibits epigenetic inheritance upon both maternal and paternal transmission. Parents with several tail kinks are more likely to produce offspring with kinked tails [17]. Interestingly, the effect is more pronounced following paternal inheritance, at least in the 129P4/RrRk mouse strain [17].

It is worth noting that all of the studies reporting A^{vy} or $Axin^{Fu}$ epigenetic inheritance use coat colour and tail morphology as phenotypic readouts of DNA methylation, respectively. No comprehensive statistical analysis has been conducted on parent and offspring DNA methylation data at the A^{vy} or $Axin^{Fu}$ loci, and very few have carried out offspring phenotyping in a manner blind to parental phenotype. While the correlation between IAP LTR methylation and phenotype severity is well documented for these classic loci, assessing the epigenetic inheritance at other metastable epialleles lacking visual phenotypes cannot rely on this form of classification. Breeding intensive experiments on six VM-IAPs quantified parental and offspring DNA methylation levels and assessed the maternal and paternal contribution to offspring VM-IAP methylation state using linear mixed-effects models. Only one region ($VM\text{-}IAP_{Gm13849}$) showed evidence of maternal inheritance and no paternal inheritance was detected (Fig. 2C). Importantly, the effect size of the maternal contribution to offspring methylation level at $VM\text{-}IAP_{Gm13849}$ was distinctly small, raising questions about its biological relevance [31].

The A^{vy} mouse model is often cited as the best described instance of mammalian epigenetic inheritance and the assumption has been that other regions in the mouse genome likely behave in the same way. The lack of heritability observed at VM-IAPs calls into question the prevalence of epigenetic inheritance across generations at metastable epialleles and warns against extrapolating from isolated examples.

4.2. Genetic background effects

The genetic background used to study the heritability of metastable epialleles is an important consideration, as both A^{vy} and $Axin^{Fu}$ inheritance patterns are influenced by the mouse strain the alleles are maintained on. The magnitude of maternal A^{vy} inheritance was shown to be dependent on whether the strain of the dam was C57BL/6J, YS/ChWf, or VY-Wf [40,42]. Similar genetic background effects have been reported for the penetrance of tail kink phenotypes associated with the $Axin^{Fu}$ allele [43]. Interestingly, when $Axin^{Fu}/+$ 129P4/RrRk male mice are crossed with A^{vy}/a C57BL/6J female mice, there is no paternal inheritance of tail phenotype, mimicking the inheritance pattern observed for A^{vy} coat colour on a C57BL/6J background [17,18].

These dependencies on genetic background suggest that the A^{vy} and $Axin^{Fu}$ parent-of-origin effects are mediated by *trans* acting genetic factors. It is possible that genetic or cytoplasmic modifiers carried in C57BL/6J oocytes, but not 129P4/RrRk oocytes, promote complete epigenetic reprogramming of metastable epialleles, therefore preventing the transmission of paternal phenotype. Such strain-specific modification has been studied in other contexts but has not been tested at metastable epialleles [44]. It is nonetheless evident that genetic background has vital implications for the design and resulting generalizability of future experiments. We are once again reminded of the interdependence of genetic and epigenetic processes, and the problematic nature of investigating epigenetic variation in genetically heterogeneous contexts.

Transposable element content and distribution differ substantially across mouse strains [45]. The C3H/HeJ strain, for example, is significantly more susceptible to new IAP insertions compared to other strains [46,47], potentially predisposing it to an increased number of metastable epialleles. As genome assemblies and the ability to map repeats improve, quality high-throughput analyses across multiple strains will be essential to understand the causes of these genetic background effects as well as the evolutionary and functional relevance of metastable epialleles.

4.3. Developmental dynamics

The mammalian genome undergoes two rounds of global epigenetic reprogramming, once during pre-implantation embryogenesis and

again during early germ cell lineage specification. After fertilization, genome-wide DNA methylation erasure occurs via active and passive demethylation of the paternal and maternal genomes, respectively. The second genome-wide demethylation event occurs only in primordial germ cells, after which sperm- and oocyte-specific methylation patterns are established (reviewed in [48,49]). These processes are important for the reacquisition of cellular totipotency in the next generation and the proper differentiation of soma and germ line.

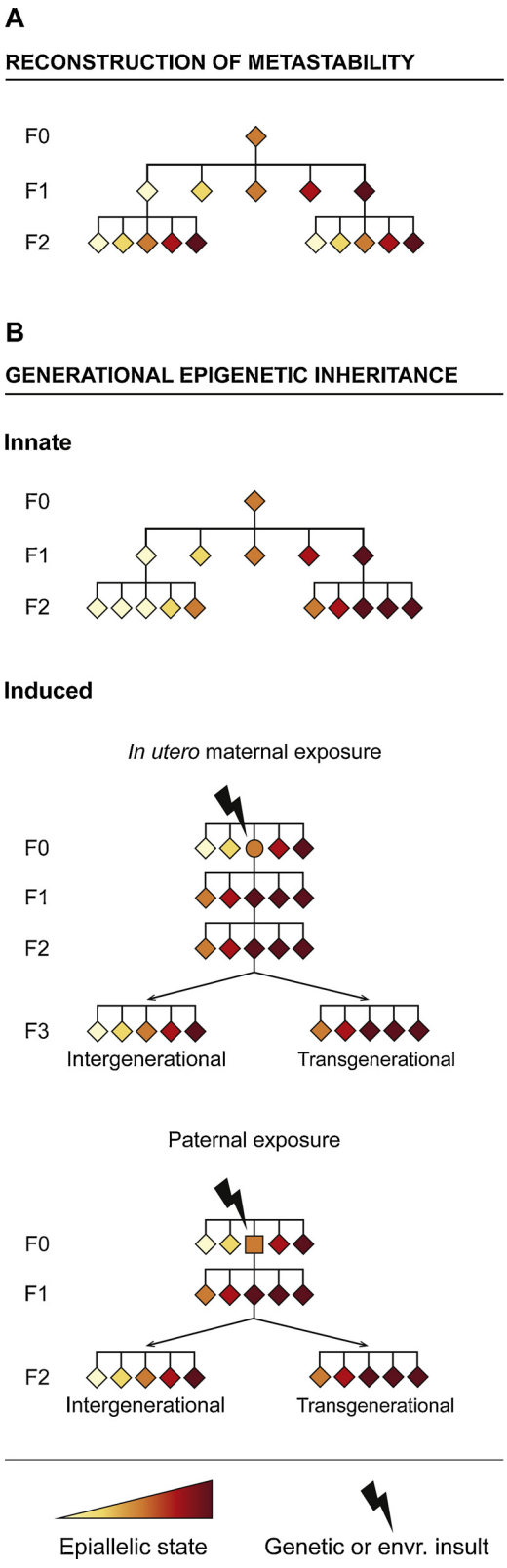
It is difficult to reconcile genome-wide reprogramming with the idea of perpetuation of epigenetic states across generations. This has led to considerable debate in the ever-growing field of epigenetic inheritance. It has been shown that some IAP elements are resistant to global demethylation in both the germline [50] and during pre-implantation development [51], providing an attractive mechanism by which the IAP-driven A^{vy} and $Axin^{Fu}$ phenotypes could be inherited. Indeed, in sperm and oocytes, the methylation levels of these alleles have been reported to reflect those observed in somatic tissues [17,52,53] (Fig. 1D). In contrast, VM-IAPs are fully methylated in mature sperm regardless of somatic methylation level [31] (Fig. 2B). Importantly, both the A^{vy} and $Axin^{Fu}$ loci are demethylated at the blastocyst stage, indicating that their associated IAPs are not resistant to the post-fertilisation wave of epigenetic reprogramming and demonstrating that DNA methylation is not perpetuated as the direct mediator of phenotypic heritability [52,53] (Fig. 1D). Little is known about the methylation dynamics at putative human metastable epialleles during early development although an analysis of human embryo methylomes suggests that the variable epigenetic states at these regions may be established during the gastrulation transition [37].

4.4. Predictable reconstruction of epigenetic stochasticity

So far our discussion on metastable epiallele epigenetic inheritance has mainly focused on a partial memory of maternal or paternal methylation levels, reflected as a phenotypic bias in the F1 generation towards parental phenotype. Notably, only three such examples have been described (A^{vy}, $Axin^{Fu}$, VM-IAP$_{Gm13849}$), and even in these cases, persistence of methylation across generations does not occur. Instead, each of these epialleles reacquires a variable methylation state in the next generation. We suggest that more careful consideration should be given to the remarkable reconstruction of this epigenetic state from one generation to the next (Fig. 3A).

While analyses of the A^{vy} and $Axin^{Fu}$ loci have largely been based on phenotypic data, VM-IAPs were identified and characterized based on their methylation profiles [31]. This created an opportunity for a more comprehensive and quantitative cross-locus assessment of methylation variability. Interestingly, VM-IAP methylation levels are highly locus-specific and are not correlated within an individual. For example, a single mouse can be highly methylated at one VM-IAP and lowly methylated at another (Fig. 2A). This suggests VM-IAPs are not uniformly targeted by the same *trans*-acting mechanism. In addition, the methylation range within a population is different from locus to locus but is remarkably constant for a given VM-IAP, even after passage through the germline and regardless of parental methylation level [31].

We suggest that it is this predictable reconstruction of epigenetic stochasticity from generation to generation that raises the most compelling mechanistic questions, rather than the subtle memory of parental methylation level observed at a minority of metastable epialleles. What is the role of genomic context in delimiting the consistent methylation ranges observed at each locus? What factors are at play during the probabilistic acquisition of methylation states within the limits of each range? At what developmental time point is parental methylation level forgotten, and when is the methylation state of offspring established? How long is the period of stochastic establishment and is there a later somatically heritable state? Which of these mechanistic aspects vary between VM-IAPs? The answers to these questions, as yet unknown, will provide considerable insight into the mechanisms underlying the resetting of epigenetic variability across generations. Once this is understood, then the basis for partial memory of parental methylation level observed for a subset of loci can be addressed.



(caption on next page)

Fig. 3. Reconstruction and heritability of epiallelic states.
(A) Reconstruction of epigenetic metastability. The full range of epiallelic states is reconstructed after passage through the germline, regardless of parental state. **(B)** Generational epigenetic inheritance occurs when the parental epiallelic state influences that of the offspring. It can either be *innate*, occurring in the absence of an external trigger, or *induced*, defined by the cross-generational persistence of an epigenetic change brought about by a genetic or environmental insult inflicted in a previous generation. In the case of induced epigenetic inheritance, if the phenotypic or epigenetic perturbation persists at least to the F3 generation following *in utero* maternal exposure (or to the F2 generation following paternal exposure), the effect is *transgenerational*. It is *intergenerational* if it only persists to F2 (or F1 upon paternal transmission) [102]. Since the distinction between trans- and intergenerational inheritance is most often associated with induced instances, it is only shown for this context above. Pedigrees: circle—female; square—male; diamond—sex unspecified.

5. Mechanistic insights into the establishment and maintenance of epigenetic metastability

5.1. Histone modifications

The inter-individual methylation variation observed at metastable epialleles is likely associated with other variable epigenetic factors. Two studies have described the presence of variable histone marks at metastable epialleles: one conducted in A^{vy} liver tissue, the other in $Axin^{Fu}$ blastocysts. Mild enrichment of H3 and H4 di-acetylation was observed at the A^{vy} IAP LTR in yellow mice while H4K20me3 enrichment was detected in pseudoagouti mice. H4K20me3 is thought to be the most prominent histone modification at IAP LTRs, targeting them specifically over other types of repeats such as L1 elements [54,55]. No difference in H3K4me3 was found [56]. The study that assessed the histone modification landscape at the $Axin^{Fu}$ locus at the blastocyst stage found significant differences in H3K4me2 and H3K9ac between blastocysts generated from penetrant and silent sires, suggesting histone marks may be involved in the transmission of tail phenotypes across generations [53]. Histone modifications have also been explored at VM-IAPs, but no consistent patterns have emerged other than an enrichment for the active marks H3K27ac, H3K9ac and H3K4me3 at the bordering regions of transcript-overlapping VM-IAPs [31]. A more in-depth characterization of metastable epiallele histone profiles, both in terms of the number of loci examined and the range of histone marks considered, will prove useful in assessing their role in establishing variable epigenetic states.

5.2. Transgene modifiers

5.2.1. Drawing parallels between metastable epialleles and transgenes

The use of transgenic mice has been vital for the study of genome function and for the modelling of human disease. One of the challenges associated with producing mice carrying exogenous DNA constructs is the often-unpredictable cell-to-cell variability in transgene expression levels within a cell population or between individuals. The cause of such variegation remains poorly understood and has been attributed to a range of factors, including the repressive effects of multi-copy transgene arrays, the proximity of the integration site to heterochromatin, and the presence of viral or plasmid-derived sequences within transgene constructs [57–61].

Reminiscent of the properties of metastable epialleles, variegated transgenes are (1) linked to variable DNA methylation levels, (2) modulated by strain background, and (3) influenced by parental origin [62–65]. For some transgenes, variable expressivity is recapitulated from one generation to the next in a predictable manner [66,67]. Others exhibit memory of parental methylation level, their silenced state stably inherited to subsequent generations after passage through the germline [65,68–70]. This heritable silencing is sometimes irreversible, and at other times reactivated upon transmission through the other parent or by crossing to a different strain.

Although fully heritable silencing is not a property of metastable epialleles, the overlapping characteristics with transgenes are worth considering while investigating the mechanisms underlying epigenetic stochasticity. Both are associated with foreign DNA sequences with regulatory potential, likely triggering similar host genome recognition and response pathways. Parallels have been drawn between transgenesis and retrotransposition before [71,72]; some have even classified variegated murine transgenes as metastable epialleles [23]. In fact, a successful screen for modifiers of variegated transgenes, described in the next section, confirms that transgenes and metastable epialleles share epigenetic modifiers.

5.2.2. MommeD mutagenesis screen

Having made key contributions to our understanding of the unique molecular behaviour of the A^{vy}, $Axin^{Fu}$ and $Cabp^{IAP}$ loci, Emma Whitelaw and her team embarked on a large-scale *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis screen for modifiers of epigenetic variability [73,74]. The study used a mouse line carrying a GFP reporter transgene expressed in a variegated fashion in red blood cells. Importantly, the variegated expression of this transgene is predictable: 55% of red blood cells express GFP in multi-generational isogenic mice. Offspring born to ENU-treated males were assessed for enhancement or suppression of variegation by screening for shifts in the percentage of GFP-expressing red blood cells. The resulting mutations were designated *Modifiers of Murine Metastable Epialleles* (*Mommes*); dominant mutations were referred to as *MommeDs*.

Mapping the mutations associated with *MommeDs,* as well as characterizing the role of the affected genes in epigenetic regulatory processes, is still ongoing [73–83]. More than 50 *MommeD* enhancers or suppressors of variegation have been identified in this screen. *MommeDs* that increase the proportion of GFP-expressing cells have mutations in genes acting as suppressors of variegation and involved in transgene silencing. Conversely, *MommeDs* resulting in a decrease in the proportion of GFP-expressing cells have mutations in genes that enhance variegation and promote transgene expression. Unsurprisingly, the majority of mutations fell into genes with known epigenetic regulatory properties. These include genes involved in DNA methylation (e.g. *Dnmt1* and *Dnmt3b)*, histone modification (e.g. *Brd1*, *Hdac1, Setdb1, Trim28*), and chromatin remodelling (e.g. *Baz1b, Pbrm1, Smarca4, Smarca5*). The full list of genes is reviewed elsewhere [72]. Many of the identified genes were previously detected in screens for modulators of position effect variegation (PEV) in *Drosophila*, reflecting the highly conserved nature of epigenetic modifiers [84].

In line with previously reported similarities between transgene and metastable epiallele epigenetic states, *MommeDs* were also found to modulate the A^{vy} locus. In particular, crossing *MommeD* heterozygotes with A^{vy}/a mice resulted in offspring with shifted coat colour distributions. In general, there was concordance between positive regulation of GFP-transgene expression and a shift in coat colour towards yellow, and vice versa. Yellow-shifting *MommeDs* include the mutants $Smchd1^{MD1}$, $Dnmt1^{MD2}$, $Setdb1^{MD13}$, and $Trim28^{MD9}$; pseudoagouti-shifting ones include $Smarca5^{MD4}$, Rlf^{MD8}, and Wiz^{MD30}. These experiments notably identified paternal effect genes, whereby wild type pups born to mutant sires exhibited changes in coat colour distribution. As a result, *Smarca5* and *Dnmt1* were the first ever reported paternal effect genes in the mouse [76]. More recently, *Setdb1* was found to exhibit similar behaviour [83].

The genes underlying *MommeDs* have diverse functions at endogenous loci extending beyond the regulation of transgene variegation. For example, previously uncharacterized *Smchd1* has been shown to regulate long-range interactions on the inactive X chromosome and at Hox genes [85,86]. $Smchd1^{MD1}$ mutants shift A^{vy} coat colour towards yellow upon maternal inheritance, but only in female offspring [73]. Interestingly, sex effects have also been observed at some VM-IAPs [31].

Complex interactions between the many genes uncovered by the *MommeD* mutagenesis screen are likely involved in the maintenance of epigenetic states at metastable epialleles. However, not all *MommeDs* have been studied with regards to their effect on A^{vy} coat colour, and perhaps due to the nature of the screen, none have been shown to affect the establishment of metastability but instead regulate its maintenance.

5.3. CTCF binding at VM-IAPs

CTCF (CCCTC binding factor) has recently emerged as a potential regulator of metastable epialleles. CTCF is a multi-functional methyl-sensitive DNA binding protein, with 41% of cell-line-specific CTCF binding sites being associated with DNA methylation at unbound loci [87–89]. The methylation-dependent sites contain CpGs at specific positions in the DNA binding site [89]. While CTCF binding at the A^{vy} and $Axin^{Fu}$ loci has not been studied, CTCF is enriched at VM-IAPs compared to their methylation invariant counterparts in multiple tissues and across different developmental time points [31]. It is plausible that a molecular antagonism between repeat element silencing via DNA methylation and the maintenance of unmethylated CTCF binding sites is contributing to the stochastic establishment of metastable epiallele methylation states [31]. An inverse relationship between VM-IAP methylation level and abundance of bound CTCF would substantiate this model. Two recent studies in humans further support an association between CTCF and epigenetic variability: one finds an enrichment for CTCF binding sites at human metastable epialleles [37] and the other implicates CTCF binding affinity in the regulation of stochastic switching between epigenetic states [39].

Despite these advances, the mechanisms underlying the establishment and maintenance of variable methylation levels at metastable epialleles remain poorly understood. Other pathways that might be involved include RNA-mediated regulation and/or a role for the recruitment of KRAB zinc finger proteins (KZFPs), which represent the largest family of transcription factors in mice and target repressive epigenetic states to retrotransposons in vertebrate genomes (reviewed in [90]). The rapid evolution displayed by KZFPs may in fact explain some of the strain-specific effects observed at metastable epialleles. The development of sequencing technologies and analytical pipelines that are optimized to include repeat elements will continue to improve our ability to address the functional and regulatory impact of these elements within and across generations.

6. Environmental modulation of metastable epialleles

6.1. Methyl supplementation in the A^{vy} mouse model

The A^{vy} mouse line has become a popular model for the study of environmentally induced epigenetic change, the most documented intervention being *in utero* methyl supplementation. Maternal dietary supplementation with methyl donors and co-factors, including folic acid, vitamin B12, choline, and anhydrous betaine, has been shown to shift A^{vy}/a offspring coat colour towards pseudoagouti [9,41,91,92] (Fig. 1C). The shift in phenotype has been attributed to an increase in methylation at the A^{vy} IAP [9]. However, having shown that the silent pseudoagouti version of the A^{vy} allele is not normally fully methylated but rather averages ~65% methylation, Cropley and colleagues compared the IAP methylation levels of methyl-exposed and unexposed pseudoagouti A^{vy}/a offspring and found no difference in methylation density at the silent IAP LTR. This suggests that the observed coat colour phenotypic change following *in utero* exposure to methyl donors is driven by an increase in methylation of the more lowly methylated allele [93]. It is possible that the increased methyl donor availability is acting indirectly via substrates other than cytosine bases at the A^{vy} allele, but this has not been tested. Consistent with this, a study in wild-derived deer mice that lack a repeat element at the *Agouti* locus showed that Agouti-controlled pelt colour is susceptible to methyl donor supplementation in the absence of a variably methylated retroelement [94].

Some of the previously discussed hallmark properties of metastable epialleles re-emerge in methyl supplementation studies. These include genetic background effects, whereby the magnitude of the A^{vy} coat colour shift is dependent on the mouse strain used for the experiment [41]. Additionally, the coat colour of offspring born to methyl-supplemented dams was found to only be altered when the A^{vy} allele was inherited paternally [92]. This is reminiscent of parent-of-origin effects observed in the absence of dietary supplementation [18]. Therefore, the fully reconstructed paternal allele may be more sensitive to modulation via methyl donor supplementation at this early embryonic stage. This does not rule out environmental sensitivity of the maternally inherited allele, since a subsequent study reported methyl supplement-induced alterations in offspring coat colour phenotypes upon maternal transmission [95].

Studies investigating environmental modulation of the epigenome often consist of exposing dams for two weeks prior to breeding followed by maintenance of the experimental regimen throughout pregnancy and lactation. While this experimental design maximizes the chances of observing an effect, it limits mechanistic inferences that would otherwise be possible by narrowing the window of exposure to a specific developmental time point. For example, the confinement of methyl supplementation to a single week during mid-gestation (corresponding to primordial germ cell migration and epigenetic reprogramming) resulted in a shift in offspring coat colour [92]. Another study on A^{vy} mice showed that feeding A^{vy}/a offspring a methyl donor diet post-weaning for a period of 29 weeks neither shifts coat colour nor IAP LTR methylation levels [96]. Together, these studies reveal that early pre-implantation embryogenesis is not the only environmentally susceptible period in development yet confirms that coat colour phenotype in A^{vy} mice and its associated epigenetic control are fixed by the age of weaning. Further experiments that fine-tune the exact period of environmental vulnerability will help identify the windows of opportunity and hence possible mechanisms contributing to changes in epigenetic state.

6.2. Innate versus induced epigenetic inheritance – a sense of semantics

Most studies on epigenetic inheritance across generations in mammals follow phenotypes or epigenetic changes triggered by ancestral exposures to environmental insults (e.g. [97–99]). Others track phenotypes in wild-type offspring caused by a mutation in a previous generation (e.g [100,101]). The volume of these studies is ever expanding. In response, it has become useful to distinguish *transgenerational* from *intergenerational* epigenetic inheritance. For true transgenerational epigenetic inheritance to take place, the induced phenotype must arise from germ cells never exposed to the original stimulus [2,3,102]. In the case of maternal exposure during pregnancy, the primordial germ cells of the developing embryo (the future F2 generation) are also exposed, so the induced change must persist at least to the F3 generation, arising from unexposed germ cells. This is not an issue for paternal exposure, so the heritable effect can be considered transgenerational if it persists to the F2 generation, and intergenerational if it does not.

This nomenclature is confusing when applied to A^{vy} mice. Many have referred to the A^{vy} mouse model as one of the best lines of evidence for transgenerational epigenetic inheritance. This is based on the pivotal finding that maternal A^{vy} phenotype, which is epigenetically controlled, influences that of the offspring [18]. As mentioned previously, grand-maternal phenotype also affects A^{vy} coat colour [18]. The extent to which this compounding effect extends beyond the F2 generation is unclear. We stress that these effects occur naturally in the population and no environmentally or genetically triggered phenotype is being tracked across generations in these experiments. If it were to be reported that the coat colour of F0 females influences that of the F3

generation, regardless of F1 and F2 coat colours, then the term *transgenerational* could be used. To our knowledge, this has not been investigated. To control for confounding F1 and F2 effects, such a study would require a large number of crosses extending down multiple generations. Therefore, it is currently unknown whether innate epigenetic inheritance at the A^{vy} locus is trans- or intergenerational.

That said, the unique non-genetic inheritance of the A^{vy} pelt patterns combined with their reported environmental susceptibilities have sparked interest in determining the heritability of environmentally induced epigenetic changes at the A^{vy} locus. *In utero* exposure to methyl donors was found to shift coat colour toward pseudoagouti in both the F1 and F2 generations without additional supplementation of F1 dams [92]. This implies that aspects of the mechanism of epigenetic change in response to exposure can persist (either directly or indirectly) throughout gamete maturation and embryo development. Whether or not the complete demethylation of A^{vy}/a embryos at the blastocyst stage occurs in this methyl-supplemented context is unknown. Given this finding, it follows that continuous methyl supplementation of F0, F1, and F2 dams might result in a cumulative phenotypic shift in offspring toward pseudoagouti. This hypothesis was tested but not substantiated [95]. However, a subsequent study showed that multi-generational methyl supplementation leads to a progressive increase in the proportion of pseudoagouti mice if the supplementation is paired with selection for the silent pseudoagouti phenotype, whereby only pseudoagouti offspring are set up for breeding to produce the next generation [103]. This cumulative effect was completely reversed after discontinuing supplementation. Despite disagreements on the merits and shortfalls of studies on this topic [104], it is clear that the effects of a maternal methyl supplementation do not extend beyond the F2 generation in the absence of continuous exposure. Hence, in the environmental (or induced) context, epigenetic inheritance at A^{vy} is intergenerational.

There is therefore a need to discriminate between innate and induced epigenetic inheritance across generations. Equally important, however, is the distinction between generational and cellular (mitotic) epigenetic inheritance to differentiate parent-to-offspring and mitotic cell-to-cell transmission, respectively, so the generational qualifier must be kept. To avoid semantic headaches, we propose reserving the use of *transgenerational* and *intergenerational* for cases where the phenotype is traced to a specific generation, and employ the more generic term *generational* otherwise. Thus, instances of generational epigenetic inheritance are innate or induced and can be further categorised as inter- or transgenerational when appropriate (Fig. 3B). Considering the direction of studies in this field, the latter distinction will most often be reserved for induced contexts. Accordingly, A^{vy} mice display innate generational epigenetic inheritance and diet-induced intergenerational epigenetic inheritance.

6.3. Additional A^{vy}-influencing environmental exposures

Other environmental insults have been found to influence coat colour in A^{vy} mice. Maternal ethanol consumption shifts offspring coat colour towards pseudoagouti, regardless of whether ethanol is administered preconceptionally or during gestation [105]. A shift in the same direction and an increase in methylation levels at the IAP LTR were observed following maternal supplementation of genistein, an isoflavone abundant in soy [106]. Intrauterine ionizing radiation has been reported to favour the silenced version of the A^{vy} allele in a dose- and sex-dependent manner, rescued by dietary anti-oxidants [107]. Maternal dietary bisphenol-A (BPA) consumption and lead exposure were independently shown to have the opposite effect on A^{vy}/a offspring coat colour distribution, shifting it towards yellow [108–110]. More recently, maternal exposure to phthalates, commonly found in plastics and cosmetics, caused altered coat colour distributions and higher IAP LTR methylation levels in A^{vy}/a offspring [111]. Finally, *in vitro* culture of zygotes to the blastocyst stage was found to significantly shift pup coat colour towards yellow and decrease IAP LTR methylation levels [112].

Research on A^{vy} environmental modulation has been controversial. A 2008 study on maternal consumption of casein and soy protein isolate, which contain genistein, showed no alteration of A^{vy}/a offspring coat colour [113]. Similarly, an extensive analysis using generalized linear mixed models on a total of 426 mouse litters and six different dietary interventions was unable to reproduce previously reported effects of BPA and genistein on A^{vy}/a offspring coat colour [114]. The same study revealed a strong parity effect, whereby changes in coat colour distribution were observed in offspring born from different parities within a single treatment group, highlighting the extreme care that must be taken in designing these experiments.

6.4. Environmental modulation of other loci

While most research programs on the environmental modulation of metastable epialleles have focused on A^{vy}, there is evidence that other loci are also susceptible. Maternal methyl supplementation causes a decrease in the incidence of kinked tails in $Axin^{Fu}/+$ offspring and an increase in DNA methylation at the $Axin^{Fu}$ locus [24]. Methylation levels at $Cabp^{IAP}$ are decreased in offspring born to BPA-exposed dams [108], and mildly increased following lead exposure [115]. As observed for A^{vy}-associated phenotypes, *in vitro* culture of $Axin^{Fu}/+$ embryos from the zygote to the blastocyst stage leads to a more severe tail kink phenotype [53].

Since the A^{vy} and $Axin^{Fu}$ loci arose from insertional mutations, commonly used laboratory mouse strains do not carry these loci. The recent identification of novel metastable epialleles in the C57BL/6J genome allows for the assessment of a repertoire of regions in the same set of environmentally perturbed mice to determine whether they respond synchronously, and to the same extent, to intrauterine environmental influences. One such study detected small tissue-specific DNA methylation differences at three variably methylated IAPs following perinatal lead exposure [110,116]. Table 1 summarises the studies conducted to date concerning the environmental modulation of metastable epialleles.

6.5. The A^{vy} mouse model: an epigenetic biosensor of environmental compromise?

The A^{vy} mouse model has been documented as a sensitive epigenetic biosensor of environmental compromise [117–119]. An ideal epigenetic biosensor is (1) particularly susceptible to a given environmental change and (2) exhibits an epigenetic response that is both predictable and easily detectable. Coat colour in A^{vy} mice appears to be acutely sensitive to slight changes in embryonic environment, likely via epigenetic influences at the A^{vy} locus. However, its innate epigenetic and phenotypic variability, established in large part by a stochastic process, is precisely what makes it a poor biological readout with little predictive value. The full range of coat colour phenotypes and associated methylation levels are observed in both control and exposed mice in these studies, requiring hundreds of mice to detect an effect and reach sufficient statistical power. Indeed, when over 2000 animals were analysed to assess the effect of *in utero* BPA and genistein exposures (both separately and together) on A^{vy} offspring coat colour, no significant shifts were observed [114]. Together, this not only makes the use of the A^{vy} mouse model a costly and inefficient biosensor of environmental perturbation, but also questions its efficacy in this context. Further studies on the more recently identified metastable epialleles in C57BL/6J mice will clarify whether some regions are better biosensors than others, or whether perhaps metastable epialleles *en masse* can potentially be used to build a multifactor epigenetic biosensor with enhanced predictive capabilities.

Table 1
Environmental modulation of metastable epialleles.

In utero exposure	Locus	Effect on offspring		Refs.
		Phenotypic shift	Methylation change	
Silencing				
Methyl donors	A^{vy}	twd. pseudoagouti	Increase	[9,41,91,92,95]
	$Axin^{Fu}$	twd. straight tail	Increase	[24]
Genistein	A^{vy}	twd. pseudoagouti	Increase	[106]
	A^{vy}	None	N/A	[114]
Casein and soy protein isolate[a]	A^{vy}	None	N/A	[113]
Ethanol	A^{vy}	twd. pseudoagouti	Increase	[105]
Ionizing radiation	A^{vy}	twd. pseudoagouti	Increase	[107]
Dibutyl phthalate	A^{vy}	twd. pseudoagouti	Increase	[111]
Activating				
BPA	A^{vy}	twd. yellow[b]	Decrease	[108]
	A^{vy}	None	N/A	[114]
	$Cabp^{IAP}$	N/A	Decrease	[108]
Lead	A^{vy}	twd. yellow	Cubic trend[c]	[110]
	$Cabp^{IAP}$	N/A	Cubic trend[c]	[110]
	IAP 110	N/A	Decrease	[116]
	IAP 236	N/A	Decrease	[116]
	IAP 506	N/A	Decrease	[116]
Embryo culture[d]	A^{vy}	twd. yellow	Decrease	[112]
	$Axin^{Fu}$	twd. kinky tail	Decrease	[53]

[a] Contains genistein.
[b] Only following high levels of exposure.
[c] Dose-dependent.
[d] Not *in utero*.

7. Functional and evolutionary relevance of metastable epialleles

The well-described relationship between IAP methylation and phenotypic outcome observed in A^{vy} and $Axin^{Fu}$ mice suggests that metastable epialleles can have a profound influence on phenotype, potentially acted upon both positively and negatively by natural selection. However, direct impact on neighbouring gene expression is not an obvious prerequisite for variable methylation at IAP elements, and in fact, inverse correlations between VM-IAP methylation and expression of nearby genes are rare [31]. This suggests that metastability *per se* is not maintained as a product of host genome hijacking of the *cis* regulatory sequences present in repeat elements. It should be noted that retroelements are capable of regulating host gene expression in a *trans* capacity in addition to their better-described *cis* regulatory potential [120]. A remarkable number of transcription factor binding sites are embedded in repeat elements; 40% of mouse CTCF binding sites are derived from transposable elements [120]. This along with the recently reported enrichment of CTCF binding at VM-IAP flanking regions suggests metastable epialleles may also play a role in orchestrating long-range regulatory networks [31]. Until these other contributions are explored, the functional relevance and evolutionary consequences of metastable epialleles remain open questions.

The metastable epialleles $Axin^{Fu}$, A^{vy} and $Cabp^{IAP}$ are associated with evolutionarily young IAPs of the IAPLTR1_Mm subclass [29,121]. VM-IAPs are similarly enriched for young IAP elements and show high levels of absence/presence polymorphism across mouse strains [31]. It is possible that inter-individual methylation variation reflects a transient epigenetic state associated with recent retrotransposition before reaching full repression. Under this premise, metastability might represent a snapshot in evolutionary time and a biologically inconsequential phenomenon. In the cases where the variably methylated transposable element positively affects host genome function, selective pressures would stabilize the epigenetic state of the element accordingly, likely in a tissue-specific manner [122].

Alternatively, locus-specific epigenetic variability itself may confer an evolutionary advantage. It has been proposed through mathematical modelling that stochastic methylation at repeat elements could allow for more rapid fixation of the element and its associated genes, as well as increase the probability of fixation in the first place [123]. Other evolutionary models have also been proposed, whereby the phenotypic plasticity conferred by metastable epialleles enables rapid adaptation to sudden environmental changes [124]. Developmental epigenetic reprogramming of these loci is consistent with this theory, allowing the re-establishment of epigenetic marks according to new environmental cues and rendering the developing embryo responsive to the environment into which it will be born. The predictable reconstruction of precise inter-individual methylation ranges observed at metastable epialleles may be a product of the controlled environments experimental mice are housed in, deliberately kept free of environmental fluctuations. The jury is still out on whether metastability is symptomatic of incomplete silencing or whether selective pressures are maintaining epigenetic stochasticity at specific regions in the genome. Further research on the now-expanded repertoire of known metastable epialleles will allow questions like these to be addressed more comprehensively.

8. Conclusion

The characterization of novel metastable epialleles in mammals has provided additional insight into their biology. It has become clear that inter-individual methylation variation at a repetitive element is not always accompanied by epigenetic inheritance or transcriptional regulation of neighbouring genes. In fact, the A^{vy} and $Axin^{Fu}$ loci appear rare in this regard. The predictable reconstruction of epigenetic variability across generations is what truly sets metastable epialleles apart from other genomic loci. Therein may lie their value, as models of biological stochasticity in the absence of genetic variation. The relatively subtle effects observed so far with respect to their environmental modulation argues against their use as biosensors of environmental change, but additional studies on other loci will help resolve this.

The advent of more sophisticated mathematical modelling approaches and the optimization of high-throughput sequencing for repeat genome analyses have been and will continue to be key in unravelling the prevalence, molecular drivers, and functional consequences of epigenetic metastability. While this review has largely focused on IAP-associated murine metastable epialleles, we anticipate research in the coming years will determine the extent to which other regions of the mouse genome are associated with variable methylation, including unique non-repetitive loci and repeat elements of other subclasses. In addition, comparative and interdisciplinary research in humans and across model organisms will enrich our understanding of the functional and evolutionary implications and mechanistic conservation of inter-individual epigenetic variation.

Acknowledgements

This work was supported by grants from the MRC (MR/R009791/1) and Wellcome Trust (WT095606) to ACFS and by PhD awards from the Cambridge Overseas Trust and the Downing College Scholarship to TMB. We are grateful to Noah Kessler and other members of the Ferguson-Smith lab for useful discussion and for valuable input during the preparation of the manuscript. We thank Jennifer Stauffer and David Schroaeder for provision of the photograph of A^{vy} mice,

Mitsuteru Ito for photo editing help, and David Walker for advice on figure design.

References

[1] L. Quadrana, V. Colot, Plant transgenerational epigenetics, Annu. Rev. Genet. 50 (2016) 467–491, https://doi.org/10.1146/annurev-genet-120215-035254.

[2] E. Heard, R.A. Martienssen, Transgenerational epigenetic inheritance: myths and mechanisms, Cell 157 (2014) 95–109, https://doi.org/10.1016/J.CELL.2014.02.045.

[3] E.A. Miska, A.C. Ferguson-smith, Transgenerational inheritance: models and mechanisms of non – DNA sequence – based inheritance, Science 354 (2016) 778–782, https://doi.org/10.1126/science.aaf4945.

[4] D.M.J. Duhl, H. Vrieling, K.A. Miller, G.L. Wolff, G.S. Barsh, Neomorphic agouti mutations in obese yellow mice, Nat. Genet. 8 (1994) 59, https://doi.org/10.1038/ng0994-59.

[5] T.J. Vasicek, L. Zeng, X.J. Guan, T. Zhang, F. Costantini, S.M. Tilghman, Two dominant mutations in the mouse fused gene are the result of transposon insertions, Genetics 147 (1997) 777–786.

[6] M.D. Cole, M. Ono, R.U.C.C. Huang, Terminally redundant sequences in cellular intracisternal A-particle genes, J. Virol. 38 (1981) 680–687.

[7] R. Rebollo, M.T. Romanish, D.L. Mager, Transposable elements: an abundant and natural source of regulatory sequences for host genes, Annu. Rev. Genet. 46 (2012) 120913153128008, , https://doi.org/10.1146/annurev-genet-110711-155621.

[8] M.M. Dickies, A new viable yellow mutation in the house mouse, J. Hered. 53 (1962) 84–86.

[9] R.A. Waterland, R.L. Jirtle, Transposable elements: targets for early nutritional effects on epigenetic gene regulation, Mol. Cell. Biol. 23 (2003) 5293–5300, https://doi.org/10.1128/MCB.23.15.5293.

[10] M.W. Miller, D.M.J. Duhl, H. Vrieling, S.P. Cordes, M.M. Ollmann, B.M. Winkes, G.S. Barsh, Cloning of the mouse agouti gene predicts a secreted protein ubiquitously expressed in mice carrying the lethal yellow mutation, Genes Dev. 7 (1993) 454–467, https://doi.org/10.1101/gad.7.3.454.

[11] T.T. Yen, A.M. Gill, L.G. Frigeri, G.S. Barsh, G.L. Wolff, Obesity, diabetes, and neoplasia in yellow A(vy)/- mice: ectopic expression of the agouti gene, FASEB J. 8 (1994) 479–488, https://doi.org/10.1096/fasebj.8.8.8181666.

[12] R.J. Miltenberger, R.L. Mynatt, J.E. Wilkinson, R.P. Woychik, The role of the agouti gene in the yellow obese syndrome, J. Nutr. 127 (1997) 1902–1907, https://doi.org/10.1093/jn/127.9.1902S.

[13] E.J. Michaud, M.J. van Vugt, S.J. Bultman, H.O. Sweet, M.T. Davisson, R.P. Woychik, Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage, Genes Dev. 8 (1994) 1463–1472, https://doi.org/10.1101/gad.8.12.1463.

[14] A.C. Argeson, K.K. Nelson, L.D. Siracusa, Molecular basis of the pleiotropic phenotype of mice carrying the hypervariable yellow (Ahvy) mutation at the agouti locus, Genetics 142 (1996) 557–567.

[15] S.C. Reed, The Inheritance and expression of fused, a new mutation in the house mouse, Genetics 22 (1937) 1–13.

[16] L. Zeng, F. Fagotto, T. Zhang, W. Hsu, T.J. Vasicek, W.L. Perry, J.J. Lee, S.M. Tilghman, B.M. Gumbiner, F. Costantini, The mouse Fused locus encodes axin, an inhibitor of the Wnt signaling pathway that regulates embryonic axis formation, Cell 90 (1997) 181–192, https://doi.org/10.1016/S0092-8674(00)80324-4.

[17] V.K. Rakyan, S. Chong, M.E. Champ, P.C. Cuthbert, H.D. Morgan, K.V.K. Luu, E. Whitelaw, Transgenerational inheritance of epigenetic states at the murine Axin (Fu) allele occurs after maternal and paternal transmission, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 2538–2543, https://doi.org/10.1073/pnas.0436776100.

[18] H.D. Morgan, H.G. Sutherland, D.I. Martin, E. Whitelaw, Epigenetic inheritance at the agouti locus in the mouse, Nat. Genet. 23 (1999) 314–318, https://doi.org/10.1038/15490.

[19] C.P. Walsh, J.R. Chaillet, T.H. Bestor, Transcription of IAP endogenous retroviruses is constrained by cytosine methylation, Nat. Genet. 20 (1998) 116–117, https://doi.org/10.1038/2413.

[20] E. Styles, R. Brink, The metastable nature of paramutable R alleles in maize. I. Heritable enhancement in level of standard Rr action, Genetics 54 (1966) 433–439.

[21] J.B. Hollick, G.I. Patterson, E.H. Coe, K.C. Cone, V.L. Chandler, Allelic interactions heritably alter the activity of a metastable maize pl allele, Genetics 141 (1995) 709–719.

[22] K. Miura, M. Agetsuma, H. Kitano, A. Yoshimura, M. Matsuoka, S.E. Jacobsen, M. Ashikari, A metastable DWARF1 epigenetic mutant affecting plant stature in rice, Proc. Natl. Acad. Sci. 106 (2009) 11218–11223, https://doi.org/10.1073/pnas.0901942106.

[23] V.K. Rakyan, M.E. Blewitt, R. Druker, J.I. Preis, E. Whitelaw, Metastable epialleles in mammals, Trends Genet. 18 (2002) 348–351, https://doi.org/10.1016/S0168-9525(02)02709-9.

[24] R.A. Waterland, D.C. Dolinoy, J.R. Lin, C.A. Smith, X. Shi, K.G. Tahiliani, Maternal methyl supplements increase offspring DNA methylation at axin fused, Genesis 44 (2006) 401–406, https://doi.org/10.1002/dvg.

[25] M.F. Lyon, Gene action in the X-chromosome of the mouse (Mus musculus L.), Nature 190 (1961) 372–373, https://doi.org/10.1038/190372a0.

[26] R. Druker, T.J. Bruxner, N.J. Lehrbach, E. Whitelaw, Complex patterns of transcription at the insertion site of a retrotransposon in the mouse, Nucleic Acids Res. 32 (2004) 5800–5808, https://doi.org/10.1093/nar/gkh914.

[27] C. Weinhouse, O.S. Anderson, T.R. Jones, J. Kim, S. a Liberman, M.S. Nahar, L.S. Rozek, R.L. Jirtle, D.C. Dolinoy, An expression microarray approach for the identification of metastable epialleles in the mouse genome, Epigenetics 6 (2011) 1105–1113, https://doi.org/10.4161/epi.6.9.17103.

[28] M.B. Ekram, K. Kang, H. Kim, J. Kim, Retrotransposons as a major source of epigenetic variations in the mammalian genome, Epigenetics 7 (2012) 370–382, https://doi.org/10.4161/epi.19462.

[29] C. Faulk, A. Barks, D.C. Dolinoy, Phylogenetic and DNA methylation analysis reveal novel regions of variable methylation in the mouse IAP class of transposons, BMC Genomics 14 (2013) 48, https://doi.org/10.1186/1471-2164-14-48.

[30] H. Oey, L. Isbel, P. Hickey, B. Ebaid, E. Whitelaw, Genetic and epigenetic variation among inbred mouse littermates: identification of inter-individual differentially methylated regions, Epigenetics Chromatin 8 (2015) 54, https://doi.org/10.1186/s13072-015-0047-z.

[31] A. Kazachenka, T.M. Bertozzi, M.K. Sjoberg-Herrera, N. Walker, J. Gardner, R. Gunning, E. Pahita, S. Adams, D. Adams, A.C. Ferguson-Smith, Identification, characterization, and heritability of murine metastable epialleles: implications for non-genetic inheritance, Cell 175 (2018) 1259–1271, https://doi.org/10.1016/J.CELL.2018.09.043 e13.

[32] Z. Wang, H. McSwiggin, S.J. Newkirk, Y. Wang, D. Oliver, C. Tang, S. Lee, S. Wang, S. Yuan, H. Zheng, P. Ye, W. An, W. Yan, Insertion of a chimeric retrotransposon sequence in mouse Axin1 locus causes metastable kinky tail phenotype, Mob. DNA 10 (2019) 17, https://doi.org/10.1186/s13100-019-0162-7.

[33] T.E. Van Baak, C. Coarfa, P.-A. Dugué, G. Fiorito, E. Laritsky, M.S. Baker, N.J. Kessler, J. Dong, J.D. Duryea, M.J. Silver, A. Saffari, A.M. Prentice, S.E. Moore, A. Ghantous, M.N. Routledge, Y.Y. Gong, Z. Herceg, P. Vineis, G. Severi, J.L. Hopper, M.C. Southey, G.G. Giles, R.L. Milne, R.A. Waterland, Epigenetic supersimilarity of monozygotic twin pairs, Genome Biol. 19 (2018) 2, https://doi.org/10.1186/s13059-017-1374-0.

[34] R.A. Waterland, R. Kellermayer, E. Laritsky, P. Rayco-Solon, R.A. Harris, M. Travisano, W. Zhang, M.S. Torskaya, J. Zhang, L. Shen, M.J. Manary, A.M. Prentice, Season of conception in rural Gambia affects DNA methylation at putative human metastable epialleles, PLoS Genet. 6 (2010) e1001252, , https://doi.org/10.1371/journal.pgen.1001252.

[35] R.A. Harris, D. Nagy-Szakal, R. Kellermayer, Human metastable epiallele candidates link to common disorders, Epigenetics 8 (2013) 157–163, https://doi.org/10.4161/epi.23438.

[36] M.J. Silver, N.J. Kessler, B.J. Hennig, P. Dominguez-Salas, E. Laritsky, M.S. Baker, C. Coarfa, H. Hernandez-Vargas, J.M. Castelino, M.N. Routledge, Y.Y. Gong, Z. Herceg, Y.S. Lee, K. Lee, S.E. Moore, A.J. Fulford, A.M. Prentice, R.A. Waterland, Independent genomewide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment, Genome Biol. 16 (2015) 1–14, https://doi.org/10.1186/s13059-015-0660-y.

[37] N.J. Kessler, R.A. Waterland, A.M. Prentice, M.J. Silver, Establishment of environmentally sensitive DNA methylation states in the very early human embryo, Sci. Adv. 4 (2018), https://doi.org/10.1126/sciadv.aat2624.

[38] P. Garg, R.S. Joshi, C. Watson, A.J. Sharp, A survey of inter-individual variation in DNA methylation identifies environmentally responsive co-regulated networks of epigenetic variation in the human genome, PLoS Genet. 14 (2018) e1007707, , https://doi.org/10.1371/journal.pgen.1007707.

[39] V. Onuchic, E. Lurie, I. Carrero, P. Pawliczek, R.Y. Patel, J. Rozowsky, T. Galeev, Z. Huang, R.C. Altshuler, Z. Zhang, R.A. Harris, C. Coarfa, L. Ashmore, J.W. Bertol, W.D. Fakhouri, F. Yu, M. Kellis, M. Gerstein, A. Milosavljevic, Allele-specific epigenome maps reveal sequence-dependent stochastic switching at regulatory loci, Science 361 (2018), https://doi.org/10.1126/science.aar3146 eaar3146.

[40] G.L. Wolff, Influence of maternal phenotype on metabolic differentiation of agouti locus mutants in the mouse, Genetics 88 (1978) 529–539.

[41] G.L. Wolff, R.L. Kodell, S.R. Moore, C.A. Cooney, Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice, FASEB J. 12 (1998) 949–957.

[42] G.L. Wolff, Genetic modification of homeostatic regulation in the mouse, Am. Nat. 105 (1971) 241–252.

[43] A.O. Ruvinsky, A.I. Agulnik, Gametic imprinting and the manifestation of the fused gene in the house mouse, Dev. Genet. 11 (1990) 263–269, https://doi.org/10.1002/dvg.1020110404.

[44] K.E. Latham, C. Sapienza, Localization of genes encoding egg modifiers of paternal genome function to mouse chromosomes one and two, Development 125 (1998) 929–935.

[45] J. Lilue, A.G. Doran, I.T. Fiddes, M. Abrudan, J. Armstrong, R. Bennett, W. Chow, J. Collins, S. Collins, A. Czechanski, P. Danecek, M. Diekhans, D.-D. Dolle, M. Dunn, R. Durbin, D. Earl, A. Ferguson-Smith, P. Flicek, J. Flint, A. Frankish, B. Fu, M. Gerstein, J. Gilbert, L. Goodstadt, J. Harrow, K. Howe, X. Ibarra-Soria, M. Kolmogorov, C.J. Lelliott, D.W. Logan, J. Loveland, C.E. Mathews, R. Mott, P. Muir, S. Nachtweide, F.C.P. Navarro, D.T. Odom, N. Park, S. Pelan, S.K. Pham, M. Quail, L. Reinholdt, L. Romoth, L. Shirley, C. Sisu, M. Sjoberg-Herrera, M. Stanke, C. Steward, M. Thomas, G. Threadgold, D. Thybert, J. Torrance, K. Wong, J. Wood, B. Yalcin, F. Yang, D.J. Adams, B. Paten, T.M. Keane, Sixteen diverse laboratory mouse reference genomes define strain-specific haplotypes and novel functional loci, Nat. Genet. 50 (11) (2018) 1574–1583, https://doi.org/10.1038/s41588-018-0223-8.

[46] I.A. Maksakova, M.T. Romanish, L. Gagnier, C.A. Dunn, L.N. Van De Lagemaat, D.L. Mager, Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line, PLoS Genet. 2 (2006) 1–10, https://doi.org/10.1371/journal.pgen.0020002.

[47] L. Gagnier, V.P. Belancio, D.L. Mager, Mouse germ line mutations due to retrotransposon insertions, Mob. DNA 10 (2019) 15, https://doi.org/10.1186/s13100-

019-0157-4.
[48] J.A. Hackett, M.A. Surani, DNA methylation dynamics during the mammalian life cycle, Philos. Trans. R. Soc. B Biol. Sci. 368 (2013), https://doi.org/10.1098/rstb.2011.0328.
[49] S.A. Smallwood, G. Kelsey, De novo DNA methylation: a germ cell perspective, Trends Genet. 28 (2012) 33–42, https://doi.org/10.1016/j.tig.2011.09.004.
[50] S. Seisenberger, S. Andrews, F. Krueger, J. Arand, J. Walter, F. Santos, C. Popp, B. Thienpont, W. Dean, W. Reik, The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells, Mol. Cell 48 (2012) 849–862, https://doi.org/10.1016/j.molcel.2012.11.001.
[51] N. Lane, W. Dean, S. Erhardt, P. Hajkova, A. Surani, J. Walter, W. Reik, Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse, Genesis 35 (2003) 88–93, https://doi.org/10.1002/gene.10168.
[52] M.E. Blewitt, N.K. Vickaryous, A. Paldi, H. Koseki, E. Whitelaw, Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice, PLoS Genet. 2 (2006) e49, https://doi.org/10.1371/journal.pgen.0020049.
[53] R. Fernandez-Gonzalez, M.A. Ramirez, E. Pericuesta, A. Calle, A. Gutierrez-Adan, Histone modifications at the blastocyst Axin1Fu locus mark the heritability of in vitro culture-induced epigenetic alterations in mice, Biol. Reprod. 83 (2010) 720–727, https://doi.org/10.1095/biolreprod.110.084715.
[54] J.H. a Martens, R.J. O'Sullivan, U. Braunschweig, S. Opravil, M. Radolf, P. Steinlein, T. Jenuwein, The profile of repeat-associated histone lysine methylation states in the mouse epigenome, EMBO J. 24 (2005) 800–812, https://doi.org/10.1038/sj.emboj.7600545.
[55] D. Rangasamy, Distinctive patterns of epigenetic marks are associated with promoter regions of mouse LINE-1 and LTR retrotransposons, Mob. DNA 4 (2013), https://doi.org/10.1186/1759-8753-4-27.
[56] D.C. Dolinoy, C. Weinhouse, T.R. Jones, L.S. Rozek, R.L. Jirtle, Variable histone modifications at the A(vy) metastable epiallele, Epigenetics 5 (2010) 637–644, https://doi.org/10.4161/epi.5.7.12892.
[57] D. Garrick, S. Fiering, D.I.K. Martin, E. Whitelaw, Repeat-induced gene silencing in mammals, Nat. Genet. 18 (1998) 56–59, https://doi.org/10.1109/ISSCC.2015.7063067.
[58] R. Festenstein, M. Tolaini, P. Corbella, C. Mamalaki, J. Parrington, M. Fox, A. Miliou, M. Jones, D. Kioussis, Locus control region function and heterochromatin-induced position effect variegation, Science 271 (1996) 1123–1125, https://doi.org/10.1126/science.271.5252.1123.
[59] K.W. Dobie, M. Lee, J.A. Fantes, E. Graham, A.J. Clark, A. Springbett, R. Lathe, M. McClenaghan, Variegated transgene expression in mouse mammary gland is determined by the transgene integration locus, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 6659–6664, https://doi.org/10.1073/PNAS.93.13.6659.
[60] J. Kjer-Nielsen, L. Holmber, K. Perera, J.D. McCluskey, Impaired expression of chimaeric major histocompatibility complex transgenes associated with plasmid sequences, Transgenic Res. 1 (1992) 182–187.
[61] D.I.K. Martin, E. Whitelaw, The vagaries of variegating transgenes, BioEssays 18 (1996) 919–923, https://doi.org/10.1002/bies.950181111.
[62] W. Reik, A. Collick, M.L. Norris, S.C. Barton, M.A. Surani, Genomic imprinting determines methylation of parental alleles in transgenic mice, Nature 328 (1987) 248–251, https://doi.org/10.1038/328248a0.
[63] C. Sapienza, A.C. Peterson, J. Rossant, R. Balling, Degree of methylation of transgenes is dependent on gamete of origin, Nature 328 (1987) 251–254, https://doi.org/10.1038/328251a0.
[64] J.L. Swain, T.A. Stewart, P. Leder, Parental legacy determines methylation and expression of an autosomal transgene: a molecular mechanism for parental imprinting, Cell 50 (1987) 719–727, https://doi.org/10.1016/0092-8674(87)90330-8.
[65] N.D. Allen, M.L. Norris, M.A. Surani, Epigenetic control of transgene expression and imprinting by genotype-specific modifiers, Cell 61 (1990) 853–861, https://doi.org/10.1016/0092-8674(90)90195-K.
[66] M.L. Opsahl, M. McClenaghan, A. Springbett, S. Reid, R. Lathe, A. Colman, C.B.A. Whitelaw, Multiple effects of genetic background on variegated transgene expression in mice, Genetics 160 (2002) 1107–1112.
[67] J.I. Preis, M. Downes, N.A. Oates, J.E.J. Rasko, E. Whitelaw, Sensitive flow cytometric analysis reveals a novel type of parent-of-origin effect in the mouse genome, Curr. Biol. 13 (2003) 955–959, https://doi.org/10.1016/S0960-9822(03)00335-X.
[68] M. Hadchouel, H. Farza, D. Simon, P. Tiollais, C. Pourcel, Maternal inhibition of hepatitis B surface antigen gene expression in trnasgenic mice correlates with de novo methylation, Nature 329 (6138) (1987) 454–456.
[69] M. Kearns, J. Preis, M. Mcdonald, C. Morris, E. Whitelaw, Complex patterns of inheritance of an imprinted murine transgene suggest incomplete germline erasure, Nucleic Acids Res. 28 (2000) 3301–3309, https://doi.org/10.1093/nar/28.17.3301.
[70] H.G.E. Sutherland, M. Kearns, H.D. Morgan, A.P. Headley, C. Morris, D.I.K. Martin, E. Whitelaw, Reactivation of heritably silenced gene expression in mice, Mamm. Genome 11 (2000) 347–355, https://doi.org/10.1007/s003350010066.
[71] R.K. Slotkin, R. Martienssen, Transposable elements and the epigenetic regulation of the genome, Nat. Rev. Genet. 8 (2007) 272–285, https://doi.org/10.1038/nrg2072.
[72] M. Blewitt, E. Whitelaw, The Use of Mouse Models to Study Epigenetics, (2013), https://doi.org/10.1101/cshperspect.a017939.
[73] M.E. Blewitt, N.K. Vickaryous, S.J. Hemley, A. Ashe, T.J. Bruxner, J.I. Preis, R. Arkell, E. Whitelaw, An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 7629–7634, https://doi.org/10.1073/pnas.0409375102.
[74] L. Daxinger, S.K. Harten, H. Oey, T. Epp, L. Isbel, E. Huang, N. Whitelaw, A. Apedaile, A. Sorolla, J. Yong, V. Bharti, J. Sutton, A. Ashe, Z. Pang, N. Wallace, D.J. Gerhardt, M.E. Blewitt, J. a Jeddeloh, E. Whitelaw, An ENU mutagenesis screen identifies novel and known genes involved in epigenetic processes in the mouse, Genome Biol. 14 (2013) R96, https://doi.org/10.1186/gb-2013-14-9-r96.
[75] L. Isbel, L. Prokopuk, H. Wu, L. Daxinger, H. Oey, A. Spurling, A.J. Lawther, M.W. Hale, E. Whitelaw, Wiz binds active promoters and CTCF-binding sites and is required for normal behaviour in the mouse, Elife 5 (2016), https://doi.org/10.7554/eLife.15082.
[76] S. Chong, N. Vickaryous, A. Ashe, N. Zamudio, N. Youngson, S. Hemley, T. Stopka, A. Skoultchi, J. Matthews, H.S. Scott, D. de Kretser, M. O'Bryan, M. Blewitt, E. Whitelaw, Modifiers of epigenetic reprogramming show paternal effects in the mouse, Nat. Genet. 39 (2007) 614–622, https://doi.org/10.1038/ng2031.
[77] A. Ashe, D.K. Morgan, N.C. Whitelaw, T.J. Bruxner, N.K. Vickaryous, L.L. Cox, N.C. Butterfield, C. Wicking, M.E. Blewitt, S.J. Wilkins, G.J. Anderson, T.C. Cox, E. Whitelaw, A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development, Genome Biol. 9 (2008) R182, https://doi.org/10.1186/gb-2008-9-12-r182.
[78] N.A. Youngson, N. Vickaryous, A. Van Der Horst, T. Epp, S. Harten, J.S. Fleming, K.K. Khanna, D.M. De Kretser, E. Whitelaw, A missense mutation in the transcription factor Foxo3a causes teratomas and oocyte abnormalities in mice, Mamm. Genome 22 (2011) 235–248, https://doi.org/10.1007/s00335-011-9317-7.
[79] L. Daxinger, H. Oey, A. Apedaile, J. Sutton, A. Ashe, E. Whitelaw, A forward genetic screen identifies eukaryotic translation initiation factor 3, subunit H (eIF3h), as an enhancer of variegation in the mouse, G3 (Bethesda) 2 (2012) 1393–1396, https://doi.org/10.1534/g3.112.004036.
[80] S.K. Harten, T.J. Bruxner, V. Bharti, M. Blewitt, T.M.T. Nguyen, E. Whitelaw, T. Epp, The first mouse mutants of D14Abb1e (Fam208a) show that it is critical for early development, Mamm. Genome 25 (2014) 293–303, https://doi.org/10.1007/s00335-014-9516-0.
[81] S.K. Harten, H. Oey, L.M. Bourke, V. Bharti, L. Isbel, L. Daxinger, P. Faou, N. Robertson, J.M. Matthews, E. Whitelaw, The recently identified modifier of murine metastable epialleles, Rearranged L-Myc Fusion, is involved in maintaining epigenetic marks at CpG island shores and enhancers, BMC Biol. 13 (2015), https://doi.org/10.1186/s12915-015-0128-2.
[82] A. Sorolla, M.R. Tallack, H. Oey, S.K. Harten, L.C. Daxinger, G.W. Magor, A.N. Combes, M. Ilsley, E. Whitelaw, A.C. Perkins, Identification of novel hypomorphic and null mutations in Klf1 derived from a genetic screen for modifiers of α-globin transgene variegation, Genomics 105 (2015) 116–122, https://doi.org/10.1016/j.ygeno.2014.09.013.
[83] L. Daxinger, H. Oey, L. Isbel, N.C. Whitelaw, N.A. Youngson, A. Spurling, K.K.D. Vonk, E. Whitelaw, Hypomethylation of ERVs in the sperm of mice haploinsufficient for the histone methyltransferase Setdb1 correlates with a paternal effect on phenotype, Nat. Publ. Gr. (2016) 1–10, https://doi.org/10.1038/srep25004.
[84] G. Schotta, A. Ebert, R. Dorn, G. Reuter, Position-effect variegation and the genetic dissection of chromatin regulation in Drosophila, Semin. Cell Dev. Biol. 14 (2003) 67–75, https://doi.org/10.1016/S1084-9521(02)00138-6.
[85] M.E. Blewitt, A.V. Gendrel, Z. Pang, D.B. Sparrow, N. Whitelaw, J.M. Craig, A. Apedaile, D.J. Hilton, S.L. Dunwoodie, N. Brockdorff, G.F. Kay, E. Whitelaw, SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation, Nat. Genet. 40 (2008) 663–669, https://doi.org/10.1038/ng.142.
[86] N. Jansz, A. Keniry, M. Trussart, H. Bildsoe, T. Beck, I.D. Tonks, A.W. Mould, P. Hickey, K. Breslin, M. Iminitoff, M.E. Ritchie, E. McGlinn, G.F. Kay, J.M. Murphy, M.E. Blewitt, Smchd1 regulates long-range chromatin interactions on the inactive X chromosome and at Hox clusters, Nat. Struct. Mol. Biol. 25 (2018) 766–777, https://doi.org/10.1038/s41594-018-0111-z.
[87] J.E. Phillips, V.G. Corces, CTCF: master weaver of the genome, Cell 137 (2009) 1194–1211, https://doi.org/10.1016/j.cell.2009.06.001.
[88] A.C. Bell, G. Felsenfeld, Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene, Nature 405 (2000) 482–485, https://doi.org/10.1038/35013100.
[89] H. Wang, M.T. Maurano, H. Qu, K.E. Varley, J. Gertz, F. Pauli, K. Lee, T. Canfield, M. Weaver, R. Sandstrom, R.E. Thurman, R. Kaul, R.M. Myers, J.A. Stamatoyannopoulos, Widespread plasticity in CTCF occupancy linked to DNA methylation, Genome Res. 22 (2012) 1680–1688, https://doi.org/10.1101/gr.136101.111.
[90] G. Ecco, M. Imbeault, D. Trono, KRAB zinc finger proteins, Development 144 (2017) 2719–2729, https://doi.org/10.1242/dev.132605.
[91] C.A. Cooney, A.A. Dave, G.L. Wolff, Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring, J. Nutr. 132 (2002) 2393S–2400S.
[92] J.E. Cropley, C.M. Suter, K.B. Beckman, D.I.K. Martin, Germ-line epigenetic modification of the murine A vy allele by nutritional supplementation, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 17308–17312, https://doi.org/10.1073/pnas.0607090103.
[93] J.E. Cropley, C.M. Suter, K.B. Beckman, D.I.K. Martin, CpG methylation of a silent controlling element in the murine A vy allele is incomplete and unresponsive to methyl donor supplementation, PLoS One 5 (2010) 1–7, https://doi.org/10.1371/journal.pone.0009055.
[94] K.R. Shorter, V. Anderson, P. Cakora, A. Owen, K. Lo, J. Crossland, A.C.H. South, M.R. Felder, P.B. Vrana, Pleiotropic effects of a methyl donor diet in a novel animal model, PLoS One 9 (2014), https://doi.org/10.1371/journal.pone.0104942.

[95] R.A. Waterland, M. Travisano, K.G. Tahiliani, Diet-induced hypermethylation at agouti viable yellow is not inherited transgenerationally through the female, FASEB J. 21 (2007) 3380–3385, https://doi.org/10.1096/fj.07-8229com.
[96] D.A. Warzak, S.A. Johnson, M.R. Ellersieck, R.M. Roberts, X. Zhang, S.M. Ho, C.S. Rosenfeld, Effects of post-weaning diet on metabolic parameters and DNA methylation status of the cryptic promoter in the A^{vy} allele of viable yellow mice, J. Nutr. Biochem. 26 (2015) 667–674, https://doi.org/10.1016/j.jnutbio.2015.01.003.
[97] B.R. Carone, L. Fauquier, N. Habib, J.M. Shea, C.E. Hart, R. Li, C. Bock, C. Li, H. Gu, P.D. Zamore, A. Meissner, Z. Weng, Ha. Hofmann, N. Friedman, O.J. Rando, Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals, Cell 143 (2010) 1084–1096, https://doi.org/10.1016/j.cell.2010.12.008.
[98] S.-F. Ng, R.C.Y. Lin, D.R. Laybutt, R. Barres, J.A. Owens, M.J. Morris, Chronic high-fat diet in fathers programs β-cell dysfunction in female rat offspring, Nature 467 (2010) 963–966, https://doi.org/10.1038/nature09491.
[99] E.J. Radford, M. Ito, H. Shi, J.A. Corish, K. Yamazawa, E. Isganaitis, S. Seisenberger, T.A. Hore, W. Reik, S. Erkek, A.H.F.M. Peters, M.-E. Patti, A.C. Ferguson-Smith, In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism, Science 345 (2014) 1255903, , https://doi.org/10.1126/science.1255903.
[100] V.R. Nelson, S.H. Spezio, J.H. Nadeau, Transgenerational genetic effects of the paternal y chromosome on daughters phenotypes, Epigenomics 2 (2010) 513–521, https://doi.org/10.2217/epi.10.26.
[101] N. Padmanabhan, D. Jia, C. Geary-Joo, X. Wu, A.C. Ferguson-Smith, E. Fung, M.C. Bieda, F.F. Snyder, R.A. Gravel, J.C. Cross, E.D. Watson, Mutation in folate metabolism causes epigenetic instability and transgenerational effects on development, Cell 155 (2013) 81–93, https://doi.org/10.1016/j.cell.2013.09.002.
[102] A.C. Ferguson-Smith, M.E. Patti, You are what your dad ate, Cell Metab. 13 (2011) 115–117, https://doi.org/10.1016/j.cmet.2011.01.011.
[103] J.E. Cropley, T.H.Y. Dang, D.I.K. Martin, C.M. Suter, The penetrance of an epigenetic trait in mice is progressively yet reversibly increased by selection and environment, Proc. R. Soc. B Biol. Sci. 279 (2012) 2347–2353, https://doi.org/10.1098/rspb.2011.2646.
[104] J.E. Cropley, C.M. Suter, D.I.K. Martin, Methyl donors change the germline epigenetic state of the Avy allele, FASEB J. 21 (2007) 3021, https://doi.org/10.1096/07-1002.
[105] N. Kaminen-Ahola, A. Ahola, M. Maga, K. Mallitt, P. Fahey, T.C. Cox, E. Whitelaw, S. Chong, Maternal ethanol consumption alters the epigenotype and the phenotype of offspring in a mouse model, PLoS Genet. 6 (2010) e1000811, https://doi.org/10.1371/journal.pgen.1000811.
[106] D.C. Dolinoy, J.R. Weidman, Ra. Waterland, R.L. Jirtle, Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome, Environ. Health Perspect. 114 (2006) 567–572, https://doi.org/10.1289/ehp.8700.
[107] A.J. Bernal, D.C. Dolinoy, D. Huang, D.A. Skaar, C. Weinhouse, R.L. Jirtle, Adaptive radiation-induced epigenetic alterations mitigated by antioxidants, FASEB J. 27 (2013) 665–671, https://doi.org/10.1096/fj.12-220350.
[108] D.C. Dolinoy, D. Huang, R.L. Jirtle, Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development, Proc. Natl. Acad. Sci. 104 (2007) 13056–13061, https://doi.org/10.1073/pnas.0703739104.
[109] O.S. Anderson, M.S. Nahar, C. Faulk, T.R. Jones, C. Liao, K. Kannan, C. Weinhouse, L.S. Rozek, D.C. Dolinoy, Epigenetic responses following maternal dietary exposure to physiologically relevant levels of bisphenol A, Environ. Mol. Mutagen. 53 (2012) 334–342, https://doi.org/10.1002/em.21692.
[110] C. Faulk, A. Barks, K. Liu, J.M. Goodrich, D.C. Dolinoy, Early-life lead exposure results in dose- and sex-specific effects on weight and epigenetic gene regulation in weanling mice, Epigenomics. 5 (2013) 487–500, https://doi.org/10.2217/epi.13.49.
[111] K. Neier, D. Cheatham, L.D. Bedrosian, D.C. Dolinoy, Perinatal exposures to phthalates and phthalate mixtures result in sex-specific effects on body weight, organ weights and intracisternal A-particle (IAP) DNA methylation in weanling mice, J. Dev. Orig. Health Dis. 10 (2019) 176–187, https://doi.org/10.1017/S2040174418000430.
[112] H.D. Morgan, X.L. Jin, A. Li, E. Whitelaw, C. O'Neill, The culture of zygotes to the blastocyst stage changes the postnatal expression of an epigentically labile allele, agouti viable yellow, in mice, Biol. Reprod. 79 (2008) 618–623, https://doi.org/10.1095/biolreprod.108.068213.
[113] T.M. Badger, M.J.J. Ronis, G. Wolff, S. Stanley, M. Ferguson, K. Shankar, P. Simpson, C.-H. Jo, Soy protein isolate reduces hepatosteatosis in yellow A^{vy}/a mice without altering coat color phenotype, Exp. Biol. Med. 233 (2008) 1242–1254, https://doi.org/10.3181/0802-RM-60.
[114] C.S. Rosenfeld, P.T. Sieli, D.A. Warzak, M.R. Ellersieck, K.A. Pennington, R.M. Roberts, Maternal exposure to bisphenol A and genistein has minimal effect on A(vy)/a offspring coat color but favors birth of agouti over nonagouti mice, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 537–542, https://doi.org/10.1073/pnas.1220230110.
[115] C. Faulk, K. Liu, A. Barks, J.M. Goodrich, D.C. Dolinoy, Longitudinal epigenetic drift in mice perinatally exposed to lead, Epigenetics 9 (2014) 934–941, https://doi.org/10.4161/epi.29024.
[116] L. Montrose, C. Faulk, J. Francis, D.C. Dolinoy, Perinatal lead (Pb) exposure results in sex and tissue-dependent adult DNA methylation alterations in murine IAP transposons, Environ. Mol. Mutagen. 58 (2017) 540–550, https://doi.org/10.1002/em.22119.
[117] R.A. Waterland, Assessing the effects of high methionine intake on DNA methylation, J. Nutr. 136 (2006) 1706S–1710S doi:0022-3166/06.
[118] D.C. Dolinoy, The agouti mouse model: an epigenetic biosensor for nutritional and environmental alterations on the fetal epigenome, Nutr. Rev. 66 (2008) S7–S11, https://doi.org/10.1111/j.1753-4887.2008.00056.x.
[119] R.L. Jirtle, The Agouti mouse: A biosensor for environmental epigenomics studies investigating the developmental origins of health and disease, Epigenomics 6 (2014) 447–450, https://doi.org/10.2217/epi.14.58.
[120] V. Sundaram, Y. Cheng, Z. Ma, D. Li, X. Xing, P. Edge, M.P. Snyder, T. Wang, Widespread contribution of transposable elements to the innovation of gene regulatory networks, Genome Res. 24 (2014) 1963–1976, https://doi.org/10.1101/gr.168872.113.
[121] C. Qin, Z. Wang, J. Shang, K. Bekkari, R. Liu, S. Pacchione, K.A. McNulty, A. Ng, J.E. Barnum, R.D. Storer, Intracisternal a particle genes: distribution in the mouse genome, active subtypes, and potential roles as species-specific mediators of susceptibility to cancer, Mol. Carcinog. 49 (2010) 54–67, https://doi.org/10.1002/mc.20576.
[122] D. Reiss, D.L. Mager, Stochastic epigenetic silencing of retrotransposons: Does stability come with age? Gene 390 (2007) 130–135, https://doi.org/10.1016/j.gene.2006.07.032.
[123] S. Branciamore, A.S. Rodin, G. Gogoshin, A.D. Riggs, Epigenetics and evolution: transposons and the stochastic epigenetic modification model, AIMS Genet. 2 (2015) 148–162, https://doi.org/10.3934/genet.2015.2.148.
[124] J. Sharif, Y. Shinkai, H. Koseki, Is there a role for endogenous retroviruses to mediate long-term adaptive phenotypic response upon environmental inputs? Philos. Trans. R. Soc. B Biol. Sci. 368 (2013), https://doi.org/10.1098/rstb.2011.0340.